The Determinants of College Graduates' Migration Decision and Its Impact on Starting Salaries in China

Yan Shi

Submitted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy
under the Executive Committee
of the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2015

ABSTRACT

The Determinants of College Graduates' Migration Decision and Its Impact on Starting Salaries in China

Yan Shi

This study examines the determinants of college graduates' work migration decision and explores its impact on college graduates' starting salaries in China, using most recent nationally representative CSLM 2011 survey data.

This study is the first one that incorporates student characteristics, institution attributes and regional characteristics (both economic and non-economic factors) in the regression analysis on determinants of work migration in the Chinese context. When investigating work migration's effects on graduates' starting salaries, in addition to the OLS model, the study also employs alternative identification strategies including instrumental variable method and propensity score matching method to account for the potential endogeneity of work migration and reduce selection bias. In addition, this study addresses sample selection issue with Heckman correction technique.

The results reveal that the following variables have a significant positive impact on college graduates' work migration: study migration, science or engineering major, student leader in high school, passed CET4, passed CET6, engineering-concentrated institution, from 985 institutions, and from 211 institutions. College graduates who possess the above individual characteristics or are from the above institutions are more likely to migrate to work. Meanwhile,

this study also finds that female, provincial GDP per capita, provincial population, provincial area size, and provincial ECI score are significant and negative determinants of work migration.

In terms of work migration's impact on college graduates' starting salaries, the weighted and unweighted OLS regression analyses reveal that new graduates who decide to migrate for work enjoy a 9.9% and 8.6% starting monthly salary premium over those graduates who do not do so, respectively. Three different PSM schemes are used to conduct PS-adjusted regressions and the results also show that work migration has a statistically significant positive impact on graduates' starting salaries. Consistent with the PSM findings, the magnitude of the coefficient estimate for work migration in IV regression is larger than the OLS estimate. Specifically, for the regression using two instrumental variables, the results find that college graduates who decide to migrate for work have a 15.8% higher starting salary compared to those who would not. These findings of this study have important implications for educational policy makers and higher education institutions in China.

Table of Contents

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGEMENTS

First I would like to express my sincere gratitude and deep admiration to my advisor in doctoral program of Economics and Education at Teachers College, Professor Mun C. Tsang, for his insightful academic guidance, continuous encouragement and strong support throughout my PhD study years. Along the way, due to some family reasons, I was physically in and out of the school at some points. It was Professor Tsang who believed in me and encouraged me to keep up and complete my dissertation. He helped me get access to the data and gave me critical and constructive comments on my dissertation project. I was a "lost child who came back to study track" and without Professor Tsang's help and patience, I could never achieved it. Professor Tsang's intelligence, profound knowledge and great kindness always inspire me. He is definitely a notable scholar and my respectful life mentor.

My great gratitude goes to my dissertation committee. I would like to thank Professor Francisco Rivera-Batiz, for his expertise in economics and education and his insightful comments and references on my study. I also sincerely thank Professor Aaron Pallas and Professor Qingmin Liu for their constructive suggestions and comments on my dissertation, especially on the methodology part. My great appreciation also goes to Professor Peter Bergman, for his devotion of precious time to my dissertation and valuable comments and suggestions on my study.

I also want to thank Professor Henry Levin, Professor Thomas Bailey, and Professor Madhabi Chatterji who expanded my knowledge in economics and education, economics of higher education, and also in the field of program evaluation. My special thanks to Professor Judith Scott-Clayton, Professor Lawrence DeCarlo and Professor Jane Monroe, I took courses

with them and they provided great training and instructions on quantitative research skills and these skills are very helpful for my dissertation research design.

I would like to especially thank Professor Jinghuan Shi and Professor Yu Zhang at Institute of Education, Tsinghua University for partnering with Center on Chinese Education to allow me to use the CSLM2011 dataset they collected, and provided me with invaluable help in accessing to and working with the data.

It was great to study with so many great fellow members in the Center on Chinese Education at Teachers College, Columbia University. I especially would like to thank Dr. Fei Guo and Dr. Li Yu. We worked together on the same CSLM dataset and encouraged each other and supported each other through this tough journey. I still remember the numerous international teleconferences we had to discuss how to clean the data and construct the index variables. I also want to thank Xin Gong, Yilin Pan, Jing Li, Haogen Yao, Dr. Henan Cheng, Professor Bin Tang and Professor Lei Zheng for their valuable discussions during my dissertation writing. I also want to thank all participants in Doctoral Seminar on Chinese Education for their valuable feedback on my dissertation. I am also indebted to my best friend Xiao Zhang, who helped me with the editing and grammar check of my dissertation.

Last but not least, my deep gratitude and love go to my loving parents, Dingshan Shi and Kangmei Lan, and my dear husband Dechao Guo. Each of them has devoted so much time and other resources to support me along the way. They are always there to trust me and encourage me. Without their selfless help and deep love to me, I could never have accomplished this dissertation. In the end, I also want to mention my two boys, Alex Fengming Guo and Eric Xuanming Guo, their happy smiles are my motives to move forward.

CHAPTER ONE:
INTRODUCTION

This chapter presents an introduction of this dissertation study. Section 1.1 introduces the context and statement of problem; Section 1.2 presents the background of the development of higher education in China, Section 1.3 provides the definition of key terms; and Section 1.4 explains the organization and structure of this dissertation.

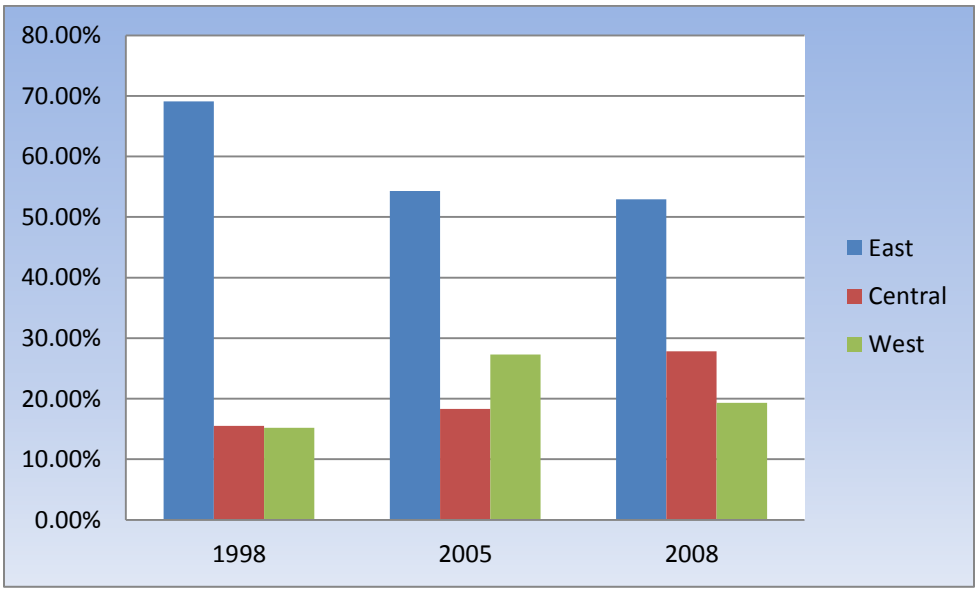
Section 1.1 Statement of Problem

Migration has been extensively discussed in labor economics and sociology for its social policy implication. The empirical investigation of college graduates' migration is of great interest to researchers internationally. Not only is such study an important component of the overall migrant stream (McHugh & Morgan, 1984), but also it closely relates to the nation's macro educational and social policy decisions. While some policy analysts regard this specific migration as a value-neutral phenomenon (WICHE, 2005), others have suggested that graduates' migration has inherent ensuing economic implications associated with the gain and loss of educated citizenry (Smith & Wall, 2006). In general, college graduates fuel a competitive workforce (Reseck et al., 2000), replenishing retiring or migrating workers, and bringing new knowledge. The lack of college graduates reduces a region's ability to build a competitive workforce (Smith & Wall, 2006).

It is commonly accepted that human capital is the primary driver of regional economic growth (Mathur, 1999; Florida, 2002; Glaeser & Saiz, 2003; Gottlieb & Fogarty, 2003). Diminishing college graduates (human capital) in a region presents an unappealing environment for local companies to flourish, because companies rely on a highly skilled and intellectual

workforce in the current economy. Especially in today's knowledge-based economy, information and knowledge are the main drives for economic development and growth. The ability to produce and use information effectively, which can be obtained through college studies, has become vital for individuals (Dean & Hunt, 2006). The most robust modern economies are those that produce the most information and knowledge, and make them easily and widely accessible (Schleicher, 2006); and this is contingent upon a highly educated citizenry (Dean & Cunningham, 2006). Therefore, preventing brain drain has been a policy imperative for many states in the U.S and worldwide.

The regional disparities of the college graduates' employment continue to be a serious problem that cannot be ignored in China. Figure 1.1 shows the regional distribution of the college graduates' employment based on findings from three different national surveys.



Data source: Zhong & Wen (2007); Ma (2010)

Figure 1-1 Regional Distribution of College Graduates' Employment in China

It's very difficult for underdeveloped regions to attract high quality college graduates with few incentives. Meanwhile these regions are losing college graduates, who move to economically more developed regions after graduation. As we know, the accumulation of human capital is crucial for social and economic development. The gap in the amount of human capital between the east and west, between urban and rural areas will widen the social and economic gap among the different regions in China. In the long run, it will jeopardize the two fundamental goals that policymakers in education want to achieve--efficiency and equity.

The Chinese government has implemented various measures to relieve the pressure of graduate unemployment and eliminate the artificial barriers to employment restrictions, such as launching the "Go West" project and reforming the residence registration system to increase mobility and integration of a national job market. The government has also initiated a number of policies (e.g. cancelling the urban expansion fees charged for receiving college graduates, simplifying the procedures to help college graduates settle down) to encourage the free flow of talent. Policy makers in a few central and western Chinese provinces have recently taken a variety of initial steps to stem the flow of college graduates outside their boundaries or to recruit new graduates from elsewhere.

In order to make efficient policies to stanch brain drain, one needs to understand what factors affect the graduates' migration decision making. Are college graduates attracted to particular regions because of economic factors such as lower living costs or lower taxes? Or do climate, scenery, culture or lifestyle options play a major role? Considerable economic research has examined how wage levels in a city or region compensate for the presence or absence of factors, such as climate, scenery, or lower living costs (Dumond, Hirsch, & MacPherson, 1999). But much of this research relies on aggregate data rather than the analysis of individual choices.

Pecuniary incentives should not be discounted, but clearly people are motivated by factors other than money (Frey, 1997). In reality, a variety of factors might interact to explain the career and location decisions of college graduates and certain factors may be more salient for some groups than others (Hansen, Ban & Huggins, 2003). Social scientists and policy makers must examine the relative importance of competing theories to develop more effective policies to reverse the brain drain and to attract more college graduates to a given region.

The consequence of work migration is also of great interest to researchers. Some researchers argued that work migration is a significant investment in human capital and will provide college graduates with monetary compensation in the future. Yet other studies have found that work migration was not related to their labor market outcomes (Chen & Coulson, 2002; Li, Zhao, & Guo, 2010). Therefore, does work migration contribute to students' post-college labor market performance? Should we encourage students to find a job in another province after graduation? These are the questions that are currently asked by China higher education's policy makers.

There were few empirical studies on college graduates' migration and its impact on the labor market outcome in China in the past; however, some attempts have emerged trying to formulate a realistically positive theory on this recently. Using the unique national College Students' Labor Market Data, this dissertation examines the determinants and subsequent labor market outcomes of college graduates' migration in China today. It develops a model on college graduates' migration in the Chinese context. The goal of this research is to bridge economic theories with empirical educational studies on this subject. Different methodologies will be employed to address the research questions, and the findings will be discussed. The new framework is likely to provide a better understanding about the ways in which personal attributes

and regional characteristics can influence college graduates' migration decisions. It seeks to inform Chinese higher education policy makers and higher education institutions in their decision to establish relevant policies in the area. Specifically, this dissertation attempts to answer two research questions:

RQ1: What are the determinants of college graduates' migration decisions in China?

RQ2: What's the impact of migration on college graduates' initial salaries?

Chapter Three will explain these research questions in detail and describe the research design for answering them.

Section 1.2 Higher Education Development in China

In this section, a brief introduction of China's higher education system and its recent development will be presented to provide background information for this dissertation. China has the largest higher-education system in the world. According to a Ministry of Education report (2012), more than 37.6 million students were enrolled in China's higher education institutions in 2012[1]. There were 9.39 million students taking the China's National College Entrance Examination (Gao Kao) in June 2014 to try to get into college. Higher education in China has played an important role in China's economic construction and social development by producing a large number of high-skill individuals and talents.

1.2.1 The Chinese higher education system

The Chinese higher education system includes various forms and offers both degree-education and non-degree education. According to Ministry of Education (MOE) statistics, there were 1,145 four-year universities and colleges in 2012.

[1] Data source: http://www.moe.gov.cn/publicfiles/business/htmlfiles/moe/s7567/list.html

In general, postsecondary education institutions can be classified into 2-3 year short-cycle colleges, the 4-year colleges, and the graduate schools. By ownership and funding source, they can be categorized as public and private. Most universities in mainland China are public institutions and are usually administered and financed by the central government, the provincial governments, or the municipal governments. Private undergraduate colleges do exist but are comparatively new; they are mostly vocational colleges sponsored by private enterprises. In 2012, there were 390 private institutions, accounting for 34% of the total number of the four-year institutions. Among these institutions, there were 303 independent colleges, which are affiliated to public universities, but remain independent in finance and administration.

The quality of higher education varies tremendously among universities. In order to improve the quality of higher education in China, the MOE launched two projects: the "211" project and the "985" project. Project 211 was launched in 1995 and was the Chinese government's new endeavor aimed at strengthening about 100 higher education institutions of higher education and key disciplinary areas as a national priority for the 21st century. By the year of 2011, there were 112 institutions in the "211" project. Project 985 is a constructive project to build world-class universities in the 21st century. In the starting phase, nine universities were included in the project. During the second phase launched in 2004, the program was expanded and 39 universities were included in this Project. These institutions are considered to be the elite institutions in China and receive additional financial support from the central government. Institutions compete with each other to get included in the project. As elite institutions receive more funding and resources and provide better education, there is severe competition among student applicants to get into these institutions.

Chinese culture values education extremely highly and most parents regard education a major means to a good life. Most college applicants are required to take the National College Entrance Examination (NCEE), which is administered at the provincial level on the same dates in June each year across the country. Applicants submit their list of preferred schools and majors[2] before or after the NCEE exam, and in most cases, college acceptances are based solely on the scores from the NCEE. In recent years several comprehensive institutions are given some extent of autonomy in admissions, which means NCEE scores are not the only standard in selecting students.

1.2.2 China's higher education expansion

China's higher education has been experiencing transformations since 1978. The most recent and influential one is the massification of higher education, which began in 1999, when the central government of China announced a policy to enlarge the scale of higher education. As a result, the number of new students admitted to college increased by over 40% in 1999 from the previous year (see Table 1-1). By 2005, the number of new college students more than quadrupled that of 1998 (Li & Xing, 2010). Meanwhile, the total number of college students in China ranked first in the world, amounting to 23 million. And the gross enrollment rate of higher education increased by 11.2 percentage points, reaching 21%. In 2009, the gross enrollment rate in higher education reached 24.2%. According to Martin Trow's (2005) definition, China has entered the stage of mass higher education, which generally has a tertiary enrollment rate of 15% to 50%.

[2] The categories of institution's major concentration include: comprehensive, science and engineering, agriculture and forestry, pharmacy and medical science, teacher training and education, language and literature, finance and economics, political science, physical education, art, ethnic minority, military and others.

Table 1-1 The Number of New Undergraduate Students Admitted Each Year into Universities and Colleges (1998-2007)

	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007
The number of new students (million)	1.08	1.6	2.21	2.68	3.21	3.65	4.2	5.05	5.46	5.66
Annual growth rate (%)	7.7	47.4	38.16	21.61	19.46	13.77	15.2	20.11	8.24	3.64

Data source: Ministry of Education statistics 1998-2008

In 2010, about 6.3 million students graduated from college in China, compared with 0.95 million in 2000 (Yue & Zhou, 2005; The Chinese Ministry of Education, 2010). With more qualified manpower joining the labor market, it is without any doubt that higher education expansion in China has made significant contribution to China's rapid economic development, even during the difficult time after Asian Financial Crisis in 1997-1998. However, the rapid expansion of higher education also raises concerns about its quality. Bai (2006) summarizes some of the obstacles to maintaining high quality in the process of higher education expansion. First, the supply of quality inputs, such as qualified faculty and infrastructure construction, cannot keep up with the rapid expansion of enrollment. Second, many three-year institutions were upgraded to four-year institutions despite of their low capability of providing adequate four-year undergraduate education. Due to the absence of effective assessment, accreditation, and qualification systems for these colleges, graduates are not adequately prepared for the job market (Zeng & Wang, 2007).

1.2.3 Employment of college graduates in China

In the era of planned economy during 1950-1985, college graduates in China were scarce resources, because there were not many college graduates every year. Students took the universal

College Entrance Exam and received free higher education, but they had no choice as to their career destinations (Bai, 2006). Under such 'Tong Zhao Tong Fen' (Graduates Employment Assignment) system, the government controlled the whole employment process of the college graduates and distributed students to different industries and regions to satisfy the skill demand of national construction projects. As result of such a system and similar wages among different regions, a high percentage of college graduates went to work in the underdeveloped central and western regions (Zhong & Wen, 2007).

The government began to charge college students tuition and fees from 1986, the previous Graduates Employment Assignment System was gradually transformed into a two-way selection system. With the establishment of socialist market economy and enactment of 'China Education Reform and Development Program' in 1993, college graduates were free to choose their own jobs and employers could hire graduates based on their merit. Graduates had been taking advantage of the changes in the job assignment system and hunted for jobs in developed areas and coastal cities (Bai, 2006). The idea of "going abroad, going to big cities and to where one can earn more money" is very popular among graduates (Li et. al., 2002).

In the late 1990s, the job market for Chinese college graduates entered difficult times. Many state-owned enterprises, which previously recruited most of the college graduates, shrank in size. The demand for college graduates from state-owned enterprises dropped as a result. On the other hand, the higher education expansion since 1999 increased the number of college graduates dramatically, i.e., the supply of college graduates rose rapidly. With the expansion of higher education, college graduates faced more and more severe competition. The economy, despite its robust growth, did not generate enough good professional jobs to absorb the influx of highly educated young adults. The number of unemployed college graduates increased by 72.4% in

2003 compared with the previous year (Yao, 2008). The problem had got worse in recent years and the increasing rate of graduate unemployment caused concerns in society (Bai, 2006). The number of college graduates who were unemployed upon graduation reached 1.5 million in 2008 (Miao & Ding, 2009) and college graduate unemployment rate increased from 18% in 2000 (Yue & Zhou, 2005) to about 27.8% in 2010 (Wu, 2010).

While there is great need for college graduates in the underdeveloped central and western regions in China, it is observed that lots of college graduates prefer to work in metropolises such as Beijing, Shanghai, Guangzhou and other cities along the southeast coast of China. The proportion of college graduates who choose to work in the central and west areas is low. According to a national survey conducted by Peking University in 2005, the employment distribution of college graduates in the eastern, central and western provinces were 54.3% 18.3% and 27.3% respectively (Zhong & Wen, 2007). The regional disparity of college graduates' employment exacerbated the graduates' unemployment problem in metropolitan cities as well. According to some news reports, there were about 6.99 million students graduated in 2013, and less than 30% of graduates in Beijing and Shanghai obtained employment offers before graduation[3]. Between 2003 and 2009, the average starting salary for China's college graduates had stayed the same while the starting salary for migrant workers during the same period rose by nearly 80% (The New York Times, 2010).

In summary, the large expansion of higher education in China has created the graduate unemployment crisis in the past decade (Zhang, Zhao, & Lei, 2012) and induces threats to the quality of higher education. In the year of 2010, the Ministry of Education issued *The Mid- and Long-term Education Reform and Development Plan (2010-2020)*, which pointed out that it's important to improve the quality of education at all levels. It also points out that the government

[3] Data source: http://news.sohu.com/20130521/n376554609.shtml

will expand vocational education to enhance student employability and create alternative employment channels at the national levels. Under this circumstance, it is necessary to examine and understand the determinants of college graduates' work migration and its impact on students' labor market performance, so that appropriate policies can be made to improve college graduates' employment prospects.

Section 1.3 Definition of Key Terms

The following are key concepts that need further explanation.

(1) Migration

Lee (1966) defined migration as a permanent or semi-permanent change of residence. However, not all kinds of spatial mobility are included in this definition. For example, temporary moves like vacation for summer, seasonal migratory workers, for whom there is no long-term residence, are excluded. No matter how short or how long, how easy or how difficult, every act of migration involves an origin, a destination, and an intervening set of obstacles.

This study will follow the definition from Tassinopoulos & Werner (1999)'s research; migration is defined as the spatial movement of labor with a simultaneous change of residence. Thus the term "migration" is associated with a permanent character.

(2) College graduates' migration

There are no clear definitions for the term "college graduates' migration" in previous literature. Based on different study purposes, college graduates' migration is defined as the movement of students who moved out of the state, county or college region after graduation. Very often the migration distance is calculated based on two zip codes. Most studies on college students' migration in China defined college graduate's migration as the college graduate's movement from the province where his/her higher institution is located to another province. To

be consistent and comparable with these studies, this study suggests that migration happens when a college graduate moves out of the province where he/she attends college and finds a job in another province, no matter whether the student's residence registration changes.

(3) College stayer, return migrant and repeat migrant

This dissertation will follow Faggian, McCann & Sheppard (2007)'s definition to classify the students into several categories according to their sequential migration behaviors to and from college. Using their definition as a reference, here are the three categories that will be used in this study. Category 1 are repeat migrants who leave their hometown region for higher education and then find their first employment in a region that is different from both their original hometown and also their higher institution locations; category 2 are return migrants, who return to find first employment near their hometown after having acquired higher education in a different region; category 3 are college stayers who find first employment in the same region where they received their higher education.

Section 1.4 Organization of the Dissertation

The remaining part of the dissertation is organized in the following manner. Chapter Two provides a review of the literature. Specifically, it first describes the general economic theories that have implications on college graduates' migration process and impact on labor market outcome. And then a summary of the empirical studies and methodologies applied on college graduates' work migration is addressed in this section. And finally the knowledge gaps will be indicated.

Chapter Three explains the research design of this dissertation study, including key research questions, conceptual framework, research methodologies, and dataset used for this empirical investigation.

Chapter Four presents the descriptive statistics on the work migration in Chinese colleges and universities; Chapter Five and Six present the empirical findings on the determinants of work migration and its impact on initial salary, respectively; and Chapter Seven summarizes the key findings and discusses the policy implications, limitations and suggestions for future studies.

CHAPTER TWO:
LITERATURE REVIEW

Section 2.1 Theories on Migration

Migration is a complex phenomenon and its analysis is proving to be increasingly interdisciplinary (Tassinopoulos & Werner, 1999). Though having received criticism from other researchers, Ravenstein (1885)'s laws of migration remain as the starting point for work in migration concept and theory. His laws are summarized below: a. Most migrants move only a short distance. b. As migrants move toward absorption centers, they leave "gaps" that are filled by migrants from more remote districts, creating migration flows that reach "the most remote corner of the kingdom." c. There is a process of dispersion, which is the inverse of absorption. d. Each migration flow produces a compensating counter-flow. e. Urban dwellers are less migratory than rural dwellers. f. Females are more migratory than males. g. Economic factors are the main cause of migration. The theoretical and empirical study of internal migration in the U.S. has a long history in economics (Greenwood, 1975, 1985). Here is the definition used for migration in this study: Spatial movement of labor with a simultaneous change of residence.

According to neo-classical equilibrium theories, wage differentials are the only reason for migration (Tassinopoulos & Werner, 1999). Workers in low-wage regions migrate to high-wage regions. The migration of workers eventually balances out the wages in the two regions after a certain time with the premises set. Thus, migrations stop when interregional wage differences no longer exist. However, this theoretical point of view does not hold when considering the actual conditions. In reality, workers are not fully aware of the true potential of their skills and abilities. Companies are not able to expect the true productivity of their future employees. As a result of this asymmetrical information, the allocation of workers and companies is inefficient. Thus an

interregional balancing out of per capita income is not to be expected (Tassinopoulos & Werner, 1999).

Since Hicks' study, most modern analyses of the decision to migrate have been based on the hypothesis that "differences in net income advantages, chiefly differences in wages, are the main causes of migration" (Hicks, 1932). Human capital theory views migration as a way of investment to improve expected future real income and employment opportunities (Mixon & Hsing, 1994). Both the individual and the whole society would obtain significant economic benefits through migration (Schultz, 1961; Becker, 1993). Migration can contribute to improving the match between workers and jobs and it is a means in promoting efficient resource allocation by shuffling workers to society's highest valued employment, therefore the whole society benefits from increased productivity.

Based on human capital theory, Sjaastad (1962) first created a framework to analyze the costs and benefits for migration as an equilibrating mechanism in a changing economy. He developed the concepts and tools to identify the important costs and returns to migration—both public and private—and also proposed methods for estimating them. He broke down the private costs into pecuniary and non-pecuniary costs. Pecuniary costs are out-of-pocket expenses of movement (increase in expenditure for food, lodging, transportation, as well as information costs, etc); while the non-pecuniary costs include foregone earnings (while individual travels, searches for, and learns about a new job, etc.) and the psychological costs of changing one's environment (Sjaastad, 1962). Private pecuniary returns consist of a positive or negative increment to the individual's real earnings obtained by moving to another place and non-pecuniary returns include locational preferences.

In summary, a person calculates a gross utility for moving to the alternative region as well as for remaining in the region of residence. The economically motivated decision to migrate depends upon whether the cost of migration exceeds the gross benefits if one migrates. The gross gain from moving may depend on personal characteristics and the cost of moving depends on observable and unobservable household characteristics which are randomly distributed among the population (Pissarides & McMaster, 1990). According to McConnell, Brue & Macpherson (2010), the present value of net income could be written as,

(2.1) $$V_p = \sum_{n=1}^{N} \frac{E_2 - E_1}{(1+i)^n} - \sum_{n=1}^{N} \frac{C}{(1+i)^n} - Z$$

V_p is the discounted present value of net income; N is the expected work time (in years) at the destination place; E_2 is the earnings obtained from the destination place at *n*th year; E_1 is the earnings gained if the individual stays in the origin place at *n*th year; *i* is the discount rate; C is the pecuniary costs of migration; Z is the psychological costs of migration. If $V_p > 0$, which means the gains to be expected in the future outweighs the benefits one gives up, therefore, the individual would choose to migrate.

If $V_p < 0$, the costs exceed the individual gains, and migration does not occur. To put it simple, migration is seen as a form of human capital investment. Individuals calculate the value of the employment opportunities available in each of the alternative labor markets, deduct the costs of making the move, and choose whichever option that maximizes the net present value of lifetime earnings.

By questioning the assumption of human capital theory—any migrant who enters the modern sector is "absorbed" into the gainfully employed at the prevailing urban real wage, Todaro (1969) brought up the absolute income difference theory to explain the migration

between rural and urban areas in developing countries. People choose to migrate from rural areas to urban areas after they consider two important variables: the urban-rural real income differential and the probability of obtaining an urban job. The probability of obtaining an urban job plays a pivotal role in the analysis.

(2.2) $$V_R(t) = \int_{t=0}^{n} Y_R(t)e^{-rt}dt$$

(2.3) $$V_u(t) = \int_{t=0}^{n} p(t)Y_u(t)e^{-rt}dt - C(0)$$

Where, $V_u(t)$ is the discounted present value of the expected urban real income stream over an unskilled worker's planning horizon; $V_R(t)$ is the discounted present value of the expected rural real income stream over the same planning horizon; $Y_R(t)$ represents net expected rural real income in period *t* based on the average real income of *x* previous periods; $Y_u(t)$ represents net urban real income in period *t*; *r* is the discount factor reflecting the degree of consumption time preference of the typical rural unskilled worker; $C(0)$ is the initial fixed costs of migration and relocation in the urban area; and $p(t)$ is the probability of having a modern sector job in period *t*.

Therefore, it's possible that the urban-rural real income differential is positive while the discounted present value of the "expected" differential is negative. The individuals would consider the probability of whether he could find a job during period t before making a decision to move. This could well explain the phenomenon that sometimes people move from the economic developed regions to seemingly "low income" region.

Rosen (1985)'s theory of equaling differences broadens the definition of income difference, which includes non-pecuniary work conditions. He again stated that people are motivated to move due to income difference and the results of migration alleviate the income difference. If

there exists income difference among people who have the same ability, either there are differences about their working conditions or there are migration costs to move to other areas. Stark & Bloom (1985) suggest that people compare their income within their group. These comparisons generate psychological costs or benefits, different feelings of relative deprivation or relative satisfaction. A person might migrate to another location to change his relative position in the same reference group or to change his reference group (Stark & Bloom, 1985). Migration decisions are often made jointly by the migrant and by some group of nonmigrants. Meanwhile many other factors contribute to the decision to migrate. Another reason why people migrate is associated with the provision of public output (education, social welfare, etc.) and amenities (parks, museums, recreation facilities, cultural institutions etc.) in the destination place.

Guided by the income-maximizing models of Hicks (1932) and Sjaastad (1962), early empirical research focused on explaining the size and direction of migration flows, as well as on determining why certain groups of individuals, such as highly educated, are more migratory. Andrew Roy's (1951) model of self-selection describes how workers sort themselves among employment opportunities, in other words, which persons find it worthwhile to migrate to the host country. He indicated that immigrants are positively selected (i.e. have above-average earnings in both the source and host countries) when the correlation between skills in the two countries is sufficiently high and when the host country has more dispersion in its earnings distribution. The immigrant population is then drawn from the upper tail of the earnings distribution because the source country "taxes" high-ability workers and "insures" less able works against poor labor market outcomes (Roy, 1951).

Borjas (1987) argues that the people who migrate are not a randomly selected sample of the population of the source regions. The migration decision is determined by a comparison of

earnings opportunities across regions, and net of migration costs (*C*). He defines the following index function:

(2.4) $$I = \log\left[\frac{w_1}{w_0+C}\right] = (\mu_1 - \mu_0 - \pi) + (\varepsilon_1 - \varepsilon_0)$$

Where *C* gives the level of migration costs, and π represents a "time-equivalent" measure of these costs ($\pi = C/w_0$). A worker migrates to the host country if $I > 0$ and remains in the source country otherwise. Migration costs *C* differs among workers. For newly arrived migrants, they have a high chance of being unemployed while they are look for work, suggesting that low-wage migrants might have higher migration costs. High-wage migrants, however, are more likely to have prior job connections and better information about job opportunities, suggesting a negative correlation between migration costs *C* and wages. It is instructive to assume initially that the time-equivalent migration costs, π, are constant in the population. The probability that a person migrates to the host region can then be written as:

(2.5) $$P = Pr\{v > (\mu_0 + \pi - \mu_1\} = 1 - \varphi(z)$$

Where $v = \varepsilon_1 - \varepsilon_0, z = (\mu_0 + \pi - \mu_1)/\sigma_v$ and φ is the standard normal distribution function. The emigration rate is negatively correlated with mean earnings in the source region and with migration costs, and is positively correlated with mean earnings in the host country. Positive selection occurs when migrants have above-average earnings in both the source and host regions, and negative selection when immigrants have below-average earnings in both region.

There are studies that attempt to explain migration patterns. Some studies (DaVanzo, 1978; Borjas, Bronars, & Trejo, 1992; Groen, 2004; Faggian, McCann & Sheppard, 2006; Gottlieb & Joseph, 2006) have used individual or household data to examine the likelihood of out-migration.

By using aggregated data, other studies (Greenwood & Hunt, 1989; Pissarides & McMaster, 1990; Gabriel, Shack-Marquez, & Wascher, 1993; Treyz, et. al., 1993; Kodrzycki, 2001; Bound et. al., 2004; Andres & Licker, 2005) have examined place-to-place migration or net migration. Studies of the determinants of migration have commonly been formulated in the context of individual utility maximization (Greenwood, 1985). And the literature has achieved some degree of consensus regarding relevant individual characteristics, labor market conditions, and non-labor market influences (Kodrzycki, Y.K., 2001).

1. Age

Age is a significant variable influencing migration and the researcher must take it into consideration when interpreting earnings differentials over space and among occupations (Sjaastad, 1962). Studies of migration repeatedly found age to play an important role. Generally speaking, the older a person is, the less likely he or she is to migrate, all else being equal. There are various reasons for this fact.

First, migration is regarded as a form of human capital investment and net gains to migration depend on age. Older migrants have fewer years to recoup their investment costs, or put it simple, older workers have a shorter period over which they can collect the returns on the migration investment (Schultz, 1961). The shorter payoff period decreases the net gains to migration, and hence lowers the probability of migration. Second, older people tend to have higher levels of human capital (work expertise) that is specific to their present employers. This human capital is not transferable to other jobs. Therefore it is not easy for them to migrate. And finally, older people often have higher migration costs than younger people; additionally, the psychic costs of migration may rise with age.

2. Distance

The probability of migration correlates inversely with the geographical proximity a person must move (Gossman et al., 1968; Lankford & Taylor, 1971; Christal, 1982; Tassinopoulos & Werner, 1999). The greater the distance to the future region of employment, the higher the expected transportation costs. Also, it is more difficult to obtain sufficient information about the destination region, which will increase the psychological costs.

Individuals who migrate to a more distant region are more likely to return to where they came from. This may be due to possibly imprecise information about the more distant region, leading to the subsequent realization that the original decision to migrate was a mistake. Moreover sociologists believe that the "cultural difference" increases in the case of a move between increasingly distant geographical areas. That is why it may seem natural for many workers from more distant home regions to limit the duration of their stay to the medium term (Tassinopoulos & Werner, 1999).

3. Gender

Many assume that men are more mobile than women. The underlying assumption is that men tend to be more attached to their careers than women, and men are therefore more likely to make the necessary moves required in order to achieve promotion (Faggian, McCann & Sheppard, 2007). On the other hand, some studies (Ravenstein, 1885; Lee, 1966; Fielding & Halford, 1993; Boyle & Halfacree, 1995) suggest that women are more mobile than men, at least across short distances. The reason for this is that women will generally have to move according to the employment locations of their male partners or spouses, particularly after women leave the workforce to rear children (Detang-Dessendre and Molho, 2000).

The early work of Long (1980), based on the 1970 Census cross-section data, suggests that the labor market experiences of migrant women in the United States differ substantially from those of men. They found that the earnings of migrant women are negatively correlated with years-since-migration.

4. Family factors

Roy's framework has been expanded to incorporate the idea that migration decisions are made in a family context (Cobb-Clark, 1990; Borjas & Bronars, 1991). The maximization of family income implies that the migrant flow contains some tied movers, persons who would not have migrated on their own but who migrate as part of the household. Also, a number of life-cycle considerations, such as marriage, entry into the labor force, start of a career are critical in a family's decision to migrate. Migrants often follow the routes previously taken by family, friends and relatives. Via multiplier effects this phenomenon can in some cases lead to unexpectedly high migratory movements of some population groups from certain regions to certain destination regions.

5. Education

Schlottmann and Herzog (1981) found educational attainment is positively related to migration. Nakosteen and Zimmer (1980, 1982)'s approach entails the estimation of separate earnings equations for migrants and nonmigrants. Using the fitted values from the earnings equation, they estimate a structural equation for the decision to migrate. The results indicated that the probability of migration increases with educational attainment and decreases with age, for self-employed persons, and for women relative to men.

6. Migration history

There is a large body of research suggesting that previous migration is highly correlated with subsequent migration. Research in the U.S. indicates once a resident migrates to another state to attend college, he or she is less likely to return upon graduation (Adelman, 2004; Kodrzycki, 2001; Perry, 2001; Tornatzkey et. al, 1998, 2001). Kodrzycki (2001) indicated that over 80% of high-school graduates who attend an in-state college continue to reside in their home state after college graduation. By contrast, only 50% of high-school graduates who attend an out-of-state college return to their home state after college graduation. Nationally, most college student migrants fail to return to and reside in their home state within 5 years of graduation. Even if a student migrant does return to their native state following college graduation, they are less likely to remain in their native state permanently.

7. Origin region and destination region's characteristics

Differential characteristics of sending and destination regions (e.g. regional size, general labor market, prevailing conditions in land and housing markets, interregional differences in both regional wages and regional employment opportunities, state and local government policies, cultural and social environment, etc.) have important impact on moving. States with higher per capita income are likely to experience more out-migration (Kyung, 1992; Abbott & Schmid, 1975). Land area of the origin state has a push effect on the student migration (Gossman et al., 1968).

Section 2.2 Empirical Studies and Methodologies Applied on College Graduates' Migration

In this section empirical findings from the educational research on the association between various factors (e.g. age, gender, parents' socioeconomic characteristics, institutional

characteristics, regional characteristics etc.) and student's migration are summarized. Then we discuss some potential methodology problems when estimating the effects of these factors and how the previous studies have dealt with these problems.

2.2.1 International empirical studies

Based on data for interstate migration in the U.S. during 1965-1970, Schlottmann and Herzog (1981) found that the probability of migration is negatively related to age, positively related to educational attainment, but has no relationship with welfare services and public outputs. In two other studies, Herzog and Schlottmann (1986, 1991) found that the migration decision is consistent with human capital theory—high-skill workers preferred lower property taxes, a quality educational environment, and more job opportunities. Based on countywide data during 1970-1980, Clark and Hunter (1992) indicated that for population aged 20-24, the net migration rates were influenced significantly by employment growth, median housing values, rainfall, educational expenditures, heating degree days, the number of theaters, the number of professional sports teams, poverty rates, statewide marginal income tax rates, countywide property taxes, and other factors. According to Kyung (1992), differences among the characteristics of origin states and destination states constitute the pull and push forces of interstate migration for individuals.

There is large quantity of literature on general human migration, however, only a handful of studies had a focus on the specific population this study will discuss--college graduates. College graduates' migration is a part of overall migration; however, there are some subtle differences from the analytical perspective. College administrators and government officials are keen to know the economics of college graduates' migration.

Greenwood (1973) estimates the magnitudes of several variables exert their influence on the geographic mobility of college graduates. Based on the data from 66 Standard Metropolitan Statistical Areas (SMSAs) in the continental United States, he uses multiple regression tools to analyze the linear effects of some factors (family income, employment, percentage change in employment between 1950 and 1960, rate of unemployment, South-non-South dummy, West-East dummy) on in-migration (out-migration, net-migration) for both the white and the nonwhite group. Greenwood recognized that the "traditional" model of migration apply well to the migration of the educated because they are likely to possess more and better information concerning alternative income and employment opportunities. The findings suggest that income and employment opportunities play an important role in the migration decision of both educated white and educated nonwhite persons. For high-income localities and for localities experiencing rapid rates of employment growth, net in-migration relative to gross migration tends to be relatively high, while for low-income localities and for localities experiencing slow rates of employment growth, net out-migration tends to be high relative to gross migration.

Yousefi & Rives (1987)'s study was based on a survey of 1,458 college graduates from Iowa's three state-supported universities. A discriminant analysis of the graduates' migration behavior showed that decisions to migrate were influenced by a number of demographic, educational and economic considerations. Being younger, male, single, without dependents, a graduate of an out-of-state high school, and an engineering major increased the likelihood of moving. Being an agriculture, veterinary medicine, or education major and being in the position of seeking a job at the time of graduation had a negative effect on the migration decision. Poor perceptions of job opportunities in Iowa increased the graduates' chances of moving after graduation. By using data from a survey of Kansas State University College of Agriculture

graduates. Barkley (1991) applied Tobit models to specify and estimate the college graduates' job mobility and job advancement. He quantified the determinants of job turnover and the number of promotions earned. Job experience was found to be the most significant determinant of labor mobility and promotion. Personal characteristics were found to have significant but small impacts on labor mobility and advancement.

Kodrzycki (2001) applied descriptive and probit regression to analyze the trend and determinants of cross-state moves of young college graduates. Based on the National Longitudinal Survey of Youth, the findings suggest that the majority of young college graduates in a state are likely to be people who went to high school or college in that state. Young adult movers are likely to be those who have moved before college. State economic and quality-of-life conditions also influence migration. The majority of moves are to states offering improved conditions along at least one dimension—high job growth, lower unemployment, higher pay, lower housing costs, or better amenities. The results imply that location preferences vary by individual, and that circumstances unobservable to researchers also help determine location decisions.

Based on the National Science Foundation's National Survey of Recent College Graduates, Tornatzky et. al. (2001) examined the migration behavior of science and engineering graduates at the master's and bachelor's levels. They suggested that the odds of an individual taking a job in-state are shown to increase more than tenfold if the individual attends college in the same state he/she goes to high school. Graduates are more likely to stay in-state if they have the following characteristics: are foreign students subsequently employed in the U.S.; majored in a field other than engineering or the physical sciences; are older than average for their class; attended a large college in a large metropolitan area; or attended college in a large state.

In an effort to understand why so many college graduates are leaving western Pennsylvania, a survey on recent college graduates from three Pittsburgh-area universities were conducted about their career and location decisions. Based on the data from the survey, Hansen, Ban & Huggins (2003) found an increase in the people who stayed in the state between 1994 and 1999. A logistic regression analysis was applied and the results showed that an improving economy, low housing costs, and ample opportunities for continuing education were the major reasons for the increase. However, the region is still losing disproportionate numbers of minorities and graduates in high-tech fields and is attracting few immigrants. Low salaries and lack of advancement opportunities, especially for women, minorities, were the primary reasons.

Gottlieb & Joseph (2006) estimated a series of random parameter logit models of the college-to-work migration decisions of technology graduates and holders of doctorates within the U.S. They included detailed information on the migration-relevant characteristics of individuals, as well as on their actual origins and destinations at the metropolitan scale. The conclusion indicated that science and technology graduates migrate to places with more educated population, other things equal; PhD graduates pay greater attention to amenity characteristics than other degree holders; and that foreign students from some immigrants groups migrate to places where those groups are concentrated.

With a focus on gender difference, Faggian, McCann & Sheppard (2007) used dichotomous, multinomial and conditional logit models to analyze the employment-migration behavior of about 380,000 U.K. university graduates. The data came from the higher education statistics agency student leavers' questionnaire. They divided migration into five types: repeat migrants, return migrants, university stayers, late migrants, non-migrants. After controlling for a range of variables related to human capital acquisition and local economic conditions, the study found out

that U.K. female graduates are generally more migratory than male graduates. The possible explanation for this is the fact that migration can be used as a partial compensation mechanism for gender bias in the labor market.

In summary, Greenwood's study used aggregated data at a national level to identify the "pull" and "push" forces of a region on migration. Some other studies (Yousefi & Rives, 1987; Hansen, Ban & Huggins, 2003) examined the reasons for migration and brought up policy implication for the "brain drain" problems at state level. There are also studies (Barkley, 1991; Gottlieb & Joseph, 2006) that explore the migration pattern for a specific group of people. The methodologies used have improved over time, including linear discriminant analysis, OLS regression, probit model, multinomial regression and logit model.

2.2.2 Empirical studies in China

There were very few empirical studies about college graduates from the perspective of their migration among regions in China. Most existing studies examined the college graduates' migration phenomenon by applying qualitative or merely descriptive methods (Ning, 2002; Lai, 2003; Zhong & Wen, 2007; Lu & Wang, 2007). The following is a brief review of empirical studies that at least applied one kind of regression tools.

Yue (2005) defined three types of college student migration. The first type is interprovincial employment, which means that students move to another province to work after graduation. In another word, the university the student attended is not in the same province where the student works for the first job (part 1 and part 2 in figure 2.1). The second type is cross-hometown employment, which indicates that students choose to work in a province other than their hometown (part 1 and part 3 in figure 2.1). The third type is called cross-hometown education,

which means that the students attended college in a province other than his hometown. The following figure shows the relationship among different migration types.

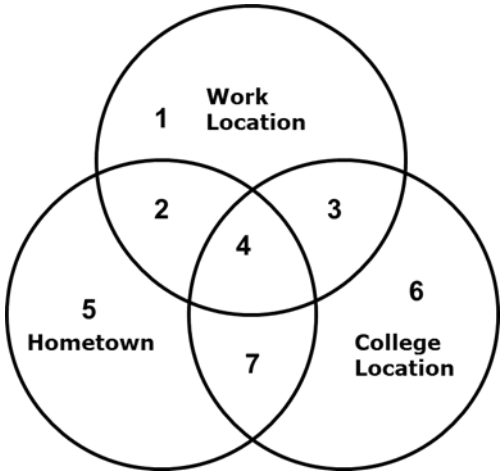


Figure 2-1 The Relationship among Work Location, Hometown, and College Location (Yue, 2005)

By using Logit regression, Yue found that there is no significant difference in job-seeking costs between inter-province employment and intra-province employment. After controlling for other factors, the initial earning of inter-province employment is much higher than that of intra-province employment. Hukou (residence registration), which brings extra costs prevents high quality graduates from taking on high-earning jobs.

Li, Liu & Guo (2009) studied the distribution of postgraduate employment migration based on a survey of postgraduates by the Chinese Academy of Science and Technology for Development (CASTED) in 2007. The survey was conducted in the cities of Beijing, Shanghai, Wuhan and Lanzhou. By using descriptive method, they found that the employment migration of the graduates are unevenly distributed, mainly concentrating in Beijing, Shanghai, Guangdong and Zhejiang yet scarcely in central and western regions.

Utilizing the data from the "2008 national survey on college graduates' employment" and "2009 national survey on college graduates' employment" done by Peking University, Ma (2010) found that students who are male, of Han ethnicity, from 985[4] & 211 colleges[5], from higher income families and whose fathers have higher education, are more likely to migrate for education. An individual's degree and historical migration are both significant factors influencing college graduates' migration choice for work. Later she applied conditional logistics model and two-stage least squares to examine the relationship between college graduates' migration and their initial earnings. She concluded that college graduates' migration helps increase their initial salary.

Based on a national-scale survey in 2009, Yue (2011) applied descriptive and regression analysis to examine college graduates' migration process. He found that about 52.4% of college graduates choose to stay in the place where they pursued college study. Graduates from coastal provinces are less likely to migrate than those from western and central regions. The level of economic development is the key factor to inter-provincial migration. Individual personal characteristics, individual human capital variables and social-economic background are main factors influencing migration. Specifically, male, minorities, and those who are not single child, who are from high quality higher institutions, who are from the countryside, are more likely to migrate. The findings also show that GDP per capita in the destination place has a significantly

[4] Project 985 was first announced at the 100th anniversary of Peking University on May 4, 1998 to promote the development of the Chinese higher education system and improve its reputation. The project involves both national and local governments allocating large amounts of funding to certain universities in order to build new research centers, improve facilities, hold international conferences, attract world-renowned faculty and visiting scholars, and help Chinese faculty attend conferences abroad (Ministry of Education of P. R. China, 2010)

[5] Project 211 was initiated in 1995 by China's Ministry of Education, with the intent of raising the research standards of high-level universities and cultivating strategies for socio-economic development (Ministry of Education of P. R. China, 2010).

positive effect on migration. The following table is a brief summary of the empirical studies on college graduates' migration in China.

Table 2-1 Empirical Studies on College Graduates' Migration in China

Study	Model	Data	Identification Strategy	Conclusion
Yue & Zhou (2005)	OLS	Survey on college graduates in 6 provinces and Beijing	No	Hukou (residence registration) brings on excess migration costs and would obstruct excellent students from getting better jobs. Evidence shows that graduates who migrated have higher initial earnings than those who didn't migrate.
Li, Zhao and Guo (2010)	Logistic Regression	Survey on postgraduates in 14 institutions in Beijing, Shanghai, Wuhan and Lanzhou cities	No	None of the other kinds of migration employment can bring significantly positive benefits in terms of higher initial earning, job matching and reduction in overeducating except those who attended a college other than his/her hometown has a higher initial earning.
Ma (2010)	OLS & Binary Logistics Model & 2SLS	2008 national survey on college graduates' employment and 2009 national survey on college graduates' employment	Yes	Students who are male, of Han ethnicity, from 985 & 211 colleges, from higher income families and whose fathers have received higher education, are more likely to migrate for education. An individual's degree and historical migration are both significant factors influencing college graduates' migration choice for work. Students who migrated have higher initial earnings.
Yue (2011)	Multinomial Logit & OLS	National-scale survey on higher institution graduates in 14 provinces in East, Central and West China	No	Graduates from coastal provinces are less likely to migrate. The level of economic development is the key factor into inter-provincial migration. Individual personal characteristics, individual human capital variables and social-economic background are main factors influencing migration. Graduates who migrated have higher initial earnings.

Li & He (2011)	Probit	Postgraduates from Beijing, Shanghai, Wuhan & Lanzhou cities	No	Both reservation wage and expected wage can improve the probability of post-graduates' migration. Being a student from big cities, from first-tier college will increase the probability of migration.

Section 2.3 Gaps in Knowledge

The existing literature on college graduates' migration in China has provided some useful insights in understanding the background and current state of the issues. However, there remain large gaps in knowledge that need to be addressed in future research. One important gap identified by the literature review is the notable lack of theoretical framework applicable in the Chinese context in the previous studies, which would inevitably undermine the strength of the studies on graduates' migration. In-depth research needs to be carried out to create solid theoretical framework for college graduates' migration in China.

Another knowledge gap is that most of the previous studies on college graduates' migration analyzed individual and location characteristics separately, focusing on either group of factors. This proposed study tries to examine the college graduates' migration comprehensively and emphasize the analysis of simultaneous interactions of key individual and place characteristics of the college graduate location choice decision. It is hoped that this approach allows for a more accurate discussion of settlement patterns of college migrants and thus might be of particular interest for policymakers at central and local levels.

Far more attention should be directed at understanding the relationship between migration and its subsequent labor market outcomes. This relationship is very important and it is directly related to the policy making. Also, whether the consequences of migration differ with individuals'

different characteristics is also an interesting topic. Currently there is no consensus on the effects of migration on college graduates' initial earnings within the Chinese context (Ma, 2010; Yue, 2011; Li, Zhao & Guo, 2010). Also it is of great interest to explore whether the determinants of migration are different by gender, or whether men are more mobile than women.

The fourth knowledge gap identified by the review is the lack of quantitative approach and rigorous econometric methodologies applied in the previous analysis in China. Due to the unavailability of appropriate data, most of the existing literature uses a qualitative or descriptive methodology, with a focus on college graduates in several specific cities. It is desirable that researchers look into the large scale quantitative approach in the future.

As indicated by Greenwood (1985), people choose to migrate because they have the rationale to believe that it will yield a higher utility than their other options. Consequently, those individuals who select a given alternative are not randomly drawn from the population as a whole. The resulting selectivity bias will make potentially serious problems in many econometric models of human behavior, including attempts to estimate the returns to migration (Greenwood, 1985). Only one study (Ma, 2010) in China made the attempt to use 2SLS (two-stage least squares) method to correct for the selection bias when discussing the effects of migration on labor market outcomes. More studies that apply rigorous methodologies are urgently needed.

CHAPTER THREE:
METHODOLOGY

In this chapter, the key research questions are brought up in section 3.1, followed by conceptual framework of this study in section 3.2. Section 3.3 presents quantitative method design for the first research question and Section 3.4 describes the quantitative method design for the second research question. Section 3.5 provides an introduction about the data that used in this study.

Section 3.1 Key Research Questions

The review of existing literature concludes that litter research has been done in finding out what factors affect college graduates' employment migration decision in China. In addition, there is no consensus on its impact. The present research's overall intent here is to study the factors underlying the college graduates' migration decision, and in this context to determine the impact of student migration on their labor market outcomes--initial salary.

This study addresses the aforementioned knowledge gaps in the following three ways: first, the study would examine both the individual and regional characteristics simultaneously, institution information will also be included; second, it is one of the first comprehensive national-scale studies that examine the graduates' migration and the impact of migration on college graduates' initial salaries; and third, this dissertation adopts some advanced quantitative methodologies that can alleviate the selection bias and other econometric concerns. Such a combination will help to identify both tangible and intangible factors that affect college graduates' migration decision making process and its consequences.

This study uses the unique College Students' Labor Market Data to address the following two key research questions by employing different analytical techniques and models:

RQ1: What are the determinants of college graduates' migration decisions in China?

RQ2: What's the impact of migration on college graduates' initial salaries?

Section 3.2 Conceptual Framework

In light of the findings from previous literature, this study develops a conceptual framework for analyzing the determinants of the college graduates' migration and its impact on labor market outcomes. Graduates' migration choices can be viewed as a function of their personal characteristics, institution attributes and regional characteristics. The graduate's desire and ability to improve his or her condition via work migration depends upon age, gender, health education, and other factors. The institution attributes include institution concentration, location and quality indicators. The "pull" of different locations depends on the strength and nature of the attraction determined by each place's economy, demography, and amenities. Furthermore, within the regional characteristics category: all the factors are grouped into demographic, economic, geographic and cultural sub-categories. Research on the above two questions is guided by the conceptual framework depicted in the figure 3.1 below.

This study is based on the theory of utility maximization. First, consider a utility maximizing college graduate *i* who selects to work in location *j*. The utility function (U_{ij}) of the individual *i* moving to *j* can be written as

(3.1) $$U_{ij} = \mathrm{V}(X_i, Z_j) + \varepsilon_{ij}$$

where X_i is a vector of personal and human capital characteristics of individual *i*, institutional attributes attached to individuals will also be included in X_i. Z_j is a vector of characteristics of the

region j, ε_{ij} is a randomly distributed error of unexplained individual, and location-specific variables. The probability $P(m_{ij})$ that an individual i will migrate to location j for employment is the probability that the individual will maximize his potential returns to human capital by entering into employment in that particular alternative region j rather than in any other region[6]. This could be written as,

$$(3.2) \qquad \mathrm{P}(m_{ij}) = prob[U_{ij} = V(X_i, Z_j) + \varepsilon_{ij} > U_{im} = V(X_i, Z_m) + \varepsilon_{im}]$$

$$\forall \mathrm{j}, \mathrm{j} \neq \mathrm{m};\ \mathrm{j}, \mathrm{m} \in \mathrm{J}$$

V_{ij} is the systematic component of utility that can be measured and ε_{ij} is the random error term.

According to Blundell & Costa-Dias (2008) and Angrist & Pischke (2008), there are several approaches to identify causal effects: (1) experiment methods, (2) natural experiment methods (i.e. differences-in-differences methods), (3) discontinuity regression (RD) methods, (4) matching methods, (5) instrumental variable (IV) methods, and (6) control function (CF) methods. All these approaches except for the first and the last one attempt to mimic the randomized assignment of the experimental setting with non-experimental data. The adaption of these identification strategies heavily depends on whether the model hypotheses are valid on specific data structure.

In addition to the identification strategies commonly used in social science, several other issues are also widely discussed in economics of education, such as omitted variables, measurement error, nonlinearity, heterogeneity, collinearity, and hierarchical data. Instrumental Variables method could address the first two problems, quantile regression (QR) could solve the

[6] Potentially work migration might be a two-stage process; however, it will not be explored here.

heterogeneous distribution problem, hierarchical linear model (HLM) and the fix/random effect model are designed to address the hierarchical data problem, and index computation could solve the collinearity problem to some extent.

Based on the conceptual framework, quantitative methodologies will be discussed in the following two sections. Considering the structure of the available data, a number of identification strategies are employed, which may improve the robustness of the results.

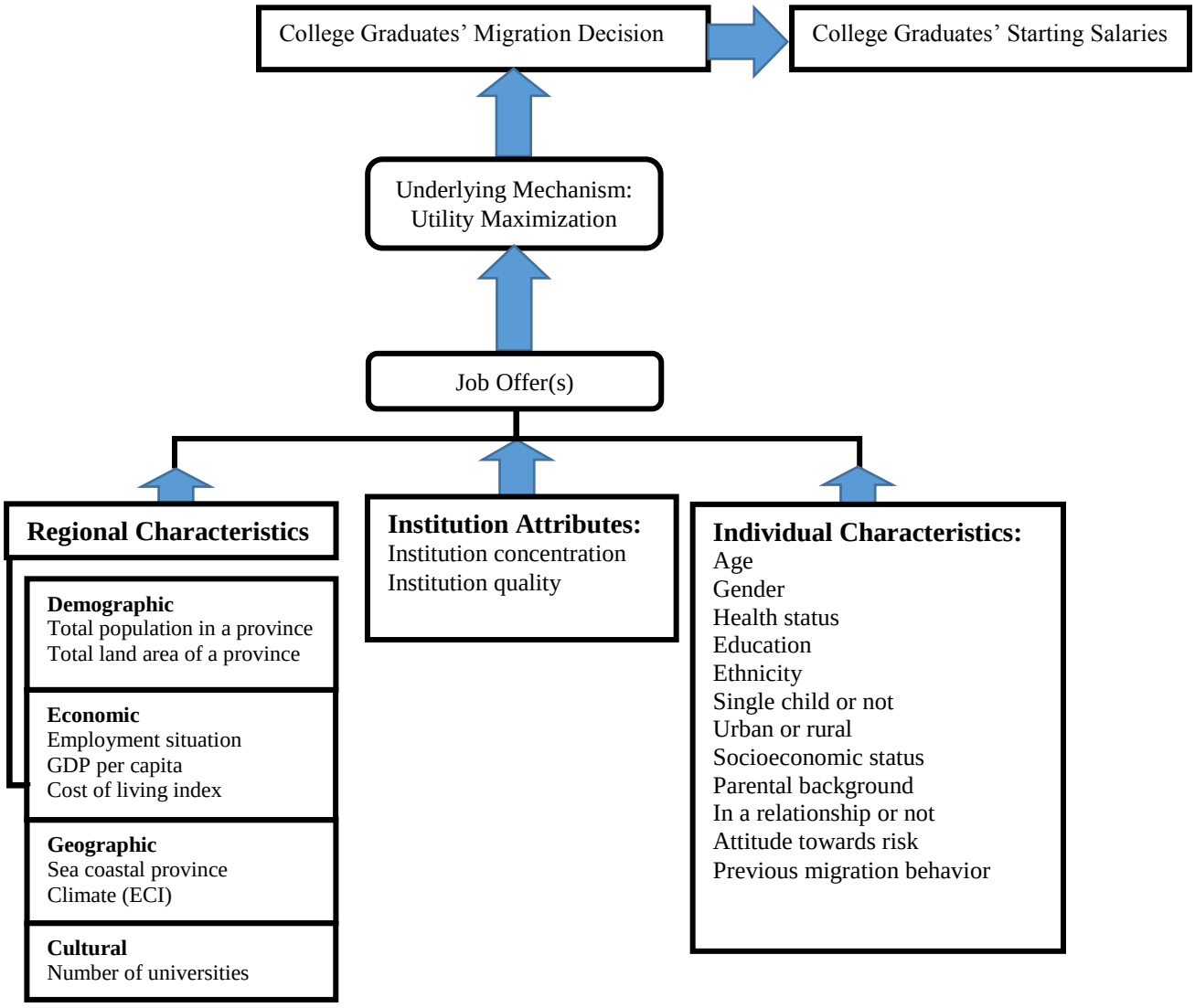


Figure 3-1 Conceptual Framework for Analysis

Section 3.3 Quantitative Designs for Estimating Determinants of Graduates' Migration

As stated in first chapter, college graduate's migration is defined as the spatial movement of a college graduate with a simultaneous change of residence. Upon graduating from higher education institution and entering first employment, a graduate must make a decision on whether he/she would move to another place. Graduates may conduct a labor market search in the region of their hometown, the region of their higher education institution, or they may seek to conduct a broader, national, labor market search.

In this study, one of the classical regression assumptions for ordinary least squares (OLS)—the dependent variable is continuous— is violated. The dependent variable is discrete, consisting of two or more outcome categories. In such circumstances, OLS poses serious inference problems and maximum likelihood techniques such as probit or logit are generally more efficient. In the expression, the unobserved portion of utility ε_{ij} is unknown. Based on different assumptions about the distribution of this unobserved utility, researchers employ different approaches to estimate the probability that an individual will chose a particular destination. In this study, probit model and multinomial logistic model will be employed to examine the migration choice problem faced by new graduates.

3.3.1 Probit model

Probit/logistic regression is used when the dependent variable is binary or dichotomous. Basically we transform the dichotomous Y into a continuous variable via a link function, known as the probit link. For binary response models, the probit and logit models are almost identical and the choice of the model is usually arbitrary. Logit models assume that the error term in the utility function follows a logistic distribution while probit model assumes the error term follows a normal distribution. First assume that there is a latent variable Y^* such that

(3.3) $$Y^* = X\beta + \varepsilon, \quad \varepsilon \sim N(0, \sigma^2)$$

In linear regression we would observe Y^* directly, however, in probit we observe only

$$y_i = \begin{cases} 0 \; if \; y_i^* \leq 0 \\ 1 \; if \; y_i^* > 0 \end{cases}$$

This translates to possible values for the error term:

(3.4) $$y_i^* > 0 \rightarrow \beta' x_i + \varepsilon_i > 0 \rightarrow \varepsilon_i > -\beta' x_i$$

$$\Pr(y_i^* > 0 | x_i) = \Pr(y_i = 1 | x_i) = \Pr(\varepsilon_i > -\beta' x_i) = \Pr(\frac{\varepsilon_i}{\sigma} > \frac{-\beta' x_i}{\sigma})$$

$$= \Phi\left(\frac{-\beta' x_i}{\sigma}\right)$$

(3.5) Similarly, $$\Pr(y_i = 0 | x_i) = 1 - \Phi\left(\frac{-\beta' x_i}{\sigma}\right)$$

In this study, when considering only two choices available (*y*=1 migrate to work; while *y*=0 do not migrate), the probability that college graduate *i* choose to migrate is given as

(3.6) $$\Pr(y_i = 1 | x_i) = \Phi\left(\frac{-\beta' x_i}{\sigma}\right)$$

When setting $\sigma = 1$, the distribution on ε is a standard normal density. In this equation, X_i includes individual-specific characteristics; and the vector of parameters β can be estimated by maximum likelihood estimation.

3.3.2 Multinomial logistic model

In the case of considering more than two choices simultaneously (repeat migrants, return migrants, college stayers), the researcher needs to distinguish the case-specific characteristics and choice-specific characteristics (Faggian, McCann, & Sheppard, 2006). Since our model is

invariant across alternatives (i.e. due to the data limitation, we only have case specific variables to work with and characteristics of the potential destination areas are not available), we use a multinomial logit model. Formally, the structure of the pure multinomial Logit model (Wooldridge, 2002) can be written as,

(3.7) $$P_{im} = Pr(y_i = m) = \frac{e^{v_{im}}}{\sum_{h=1}^{m} e^{v_{ij}}} = \frac{e^{\beta_m X_i}}{\sum_h e^{\beta_h X_i}}, \quad m = 1, \dots, 3$$

Where P_{im} is the probability that individual i chooses migration behavior m = {college stayer, return migrant, repeat migrant} and X denotes the regressor matrix. Here m =3 indicating the three different categories of sequential migration behavior exhibited by individuals. X_i is the vector of case-specific characteristics, and β_m are the parameters to be estimated by maximizing the log-likelihood function. The vector of β_m is attached to the vector of case-specific characteristics X_i that influence utility.

Based on utility maximization theory, multinomial logistic model treats college graduate's decision (discrete choice situation) as a comparison among the utilities (continuous latent variable) of alternative migration types. The multinomial logit model can be considered as simultaneously estimating binary logit models for all comparisons among the dependent categories. When using Stata to conduct the estimation, we can select the base category for comparison. The results from Stata then reports coefficients for the effect of each independent variable on each category relative to the base category.

The multinomial logic model is a popular framework to estimate the determinants of location choice for immigrants and migrants (Scott, Coomes & Izyumov, 2005). Multinomial logistic regression does not assume normality, linearity, or homoscedasticity; however, it does have a strong assumption. It assumes the independence of irrelevant alternatives (IIA), which

states that the relative probabilities of two options being chosen are unaffected by introduction or removal of other alternatives. In other words, alternatives are assumed to be independent from one another. Adding or deleting outcomes does not affect the odds among the remaining outcomes. This property suggests that alternatives are inherently unique due to factors both observed by the analyst and factors that are unobserved. More importantly, they are viewed as unique by decision makes. If the IIA property is violated then multinomial logit regression results will be biased, and hence a discrete choice model that does not require the IIA property should be used. Three common approaches to bypass the IIA restriction are nested logit, multinomial probit and mixed logit models. The mixed logit approach is regarded as the most general and flexible of the three (Hausman and Wise, 1978; Train, 2003).

There are two tests of the IIA assumption. Hausman and McFadden (1984) proposed a Hausman test and McFadden, Tye, and Train (1976) proposed an approximate likelihood-ratio test that was improved by Small and Hsiao (1985). For both the tests, multiple tests of IIA are possible. Assuming that the multinomial logit model is estimated with base category a, J-1 tests can be computed by excluding each of the remaining categories to form the restricted model. The results of the test differ, depending on which base category was used to fit the model. To test whether the multinomial model is appropriate, the Hausman and McFadden (1984) test for the IIA property was employed in this study.

The following Table 3-1 shows the definition and measures of the variables that will be included in the probit and multinomial models in this dissertation. The covariates set (X_i) in these models are almost the same. In general, it includes individual characteristics, institution attributes, and regional characteristics.

Individual characteristics include students' previous migration behavior, gender, age, race, health status, relationship, risk appetite, only child, whether the student has a rural registration of residence, academic track, socio-economic status (SES) index, National College Entrance Examination (NCEE) score, average course score in college, major, English proficiency, student leader in college, Chinese Communist Party (CCP) member, has professional certificates. Based on findings from literature review, here we included study migration as a measure to indicate students' previous migration behavior. Whether the student is in good health might affect his/her decision of work migration because graduates who are unhealthy tend to move back to hometown for a more comfortable and familiar environment. Whether the student is the only child in one's family is a measure of family structure. Those who are the only child in their family usually receive more support from the family and may choose to work in hometown. Socio-economic status index is constructed based on annual household income, type and area of resident dwelling, parents' years of schooling, and parents' occupations. Details about the construction of this index are explained in Section 3.5.1.

NCEE score is used as a measure of academic ability in previous studies, as the exam is designed to sort students into different levels of higher education institutions (Li, Meng, Shi, & Wu, 2012). Student's academic track in high school was included as a control for systematic difference in NCEE scores between different tracks. Students in humanities track use different versions of exam papers from science track students and normally they have lower NCEE scores than science-track students. Similarly, average course score in college is also included as a measure of academic ability.

Whether the student was a student leader in high school is included in the regression to control for pre-college variance in their non-cognitive abilities. As suggested by previous

literature (Schneider & Paul, 1999; Lu, 2008; Gottfried et al., 2011; Li & Lang, 2012), students who were leaders in high school possess better inter-personal skills, problem-solving skills, and they have positive motivation and attitude about future. These are the factors that may influence students' development and achievement in and after college.

Average course score in college is used as a measure of academic ability. Academic major may have an impact on graduates' migration due to unbalanced industry distribution in China. It is also an important predictor of labor market outcomes because it determines graduates' occupation and industry in the labor market. The preference degree towards one's major is an ordinal variable indicating self-reported degree to which a student likes his/her major, with 1 being "do not like at all" to 4 being "like it very much". It is a measure of students' attitude and motivation. The hypothesis is that students with higher degree of preference towards their major are more motivated and therefore may make effort to find better jobs and achieved better labor market performance.

English proficiency is included in the models for migration because it is one of the common credentials required by employers. It is measured by whether the student passed the level 4 or level 6 in the CET test. Extra-curricular experience is captured by whether the student is a member of the Chinese Communist Party (CCP), whether the student was a leader in department or institution student organizations, and whether the student has professional certificates. These variables, including performance in CET tests, are common covariates included in previous studies on post-college labor market performance in China (Yan & Mao, 2008, Du & Yue, 2010; Guo, et al., 2010; Lai et al., 2012; Xie & Li, 2010).

Institution characteristics include region of the institution, campus location of the institution, academic ranking level of the institution, academic concentration of the institution. Region of

institution is a categorical variable representing five regions in China[7]: municipalities (Beijing, Shanghai, and Tianjin), east area, northeast area, central area, and west area. The east and northeast regions are more developed and therefore have more job opportunities. The three municipalities are cities directly governed by the central government. Though located in the east region, they are listed as a separate category because there are more educational resources and job opportunities in these cities than in other places. The fourth municipality Chongqing, which is located in west China, is not included in this category. It has the shortest history of being a municipality and is far behind the development level of the other three municipalities.

Campus location is a categorical variable indicating whether the institution campus is in urban area of large cities, in both urban and suburban areas of large cities, in suburban area of large cities, or in a small city. Here large or small city is determined by the administrative level and population of the city. Small cities refer to cities at the prefecture level or below and have a population less than two million. Institutions in these cities are all located in urban area; however, as the cities are small, there may be fewer job opportunities for college graduates. As for institutions in large cities, many of them have built up new campuses in suburban area since the expansion of higher education in China. Some institutions place all undergraduate students in the suburban campus throughout their college years, while others place freshmen and sophomores in suburban campus and senior students in urban campus. Some institutions allocate students in urban and suburban campus based on academic departments.

Academic ranking of institutions refers to elite college (985 institution and 211 institution), non-key college, and independent institution. Institutions in different levels have different amount of educational resources, and therefore are different in education quality. Academic

[7] Another reason to use this 5-category region variable is that these categories were employed as one of the criteria to select participating institutions in the data collection process.

concentration of an institution is a three-category variable indicating whether the institution is a comprehensive institution, a science and engineering concentrated institution, or an institution with other concentrations. Institutions with different concentration may have different institution characteristics, for instance, composition of students, overall climate, and aim and purpose of education. These factors may influence students' college experience and outcomes.

In addition to above covariates, another set of variables measuring regional characteristics is included in the models: provincial land area size, provincial population, unemployment rate, gross domestic product (GDP) per capita, Consumer Price Index (CPI), ecological civilization index, number of higher education institutions, sea coastal province dummy. The Ecological Civilization Index is a measure of the extent that the natural environment of an area is contaminated, and its impact on human health status. As a new mode of civilization that reflects a level of harmony between humans and nature, ecological civilization represents a major conceptual advance for the development of human civilization. The higher the ECI, the more developed human and nature civilization is. In detail, the ECI consists of two parts: the Ecological Efficiency Index (Eco-efficiency index or EEI) and the Environmental Quality Index (EQI). The EEI and EQI were weighted and then calculated to become the 2013 revised ECI (Liu, 2014). The EEI measures the degree and efficiency of the region's ecological resources consumption to achieve economic development in the region. Gross domestic product (GDP) being equal, the smaller the impact of economic development on the natural environment, the higher the EEI is. The EQI characterizes the quality of the living environment from its air quality point of view and is directly related to the quality of people's lives. To calculate the EQI, the air quality index (Air Quality Integrated Index, AQII) was adopted as the core indicator of the environmental quality of a region.

Table 3-1 Definition and Measures of Variables Used in the Probit and Multinomial Models

Variable Name	Definition	Core Measurement or Comments
Dependent Variables:		
Work migration decision	Graduates' migration decision: migrate to work; do not migrate to work	Dummy variable: 1=migrate to work, 0=do not migrate to work
Work migration status	Graduates' migration status: college stayers, return migrants, repeat migrants	Categorical variable: 1=college stayer, 2=return migrant, 3=repeat migrant
Covariates:		
Individual Characteristics		
Study migration	Whether or not a student migrated to attend college	Dummy variable: 1=migrated to attend college, 0=did not migrate to attend college
Female	Gender of the student	Dummy variable: 1=female, 0=male
Age	Age of the student in 2011	Continuous variable, calculated by birthday question
Health status	Self-reported health status of the student	Dummy variable: 1=unhealthy, 0=healthy
Minority	Whether the student is from a minority ethnic group	Dummy variable: 1=minority, 0=Han
From rural area	Whether the student has a rural residence registration (Hukou)	Dummy variable: 1=rural, 0=urban
In a relationship	Whether the student reported in a relationship	Dummy variable: 1=yes, 0=no
Risk appetite	Student's attitude towards risk	Categorical variable: 1=risk aversive, 2=risk neutral, 3=risk seeking
Only child	Whether the student is the only child in their family	Dummy variable: 1=only child, 0=not only child)
SES score	Constructed index of the socio-economic status of the student's family	Continuous variable; it's a composite score based on parents' years of education, parents' occupations, annual household income, and family wealth measured by assets

Student leader in high school	Whether the student was a leader in high school	Dummy variable: 1=yes, 0=no
Humanities track in high school	Whether the student was on humanities academic track in high school	Dummy variable: 1=yes, 0=no
Arts or sports track in high school	Whether the student was on arts or sports track in high school	Dummy variable: 1=yes, 0=no
NCEE score	National college entrance examination score (rescaled to 1-100), measure of academic ability	Continuous variable
Science or engineering major	Whether the student has a science or engineering major	Dummy variable: 1=yes, 0=no
Economics or management major	Whether the student has an economics or management major	Dummy variable: 1=yes, 0=no
Average course score in college	Student's average score in college	Continuous variable
Preference degree towards one's major	How the student likes his/her major	Ordered categorical variable: 1=not at all, 2= a little bit, 3=somewhat, 4=very much
Passed CET4	Whether the student passed CET4 test (English proficiency measure)	Dummy variable: 1=yes, 0=no
Passed CET6	Whether the student passed CET6 test (English proficiency measure)	Dummy variable: 1=yes, 0=no
Student leader in college	Whether the student was a leader in college	Dummy variable: 1=yes, 0=no
CCP member	Whether the student is CCP member	Dummy variable: 1=yes, 0=no
Has professional certificates	Whether the student has professional certificates	Dummy variable: 1=yes, 0=no
Institution Attributes		
Institution region	Region of the institution	Categorical variable: 1=municipalities, 2=east area, 3=northeast area, 4=central area, 5=west area
Institution concentration	Concentration/Specialization of the institution	Categorical variable: 1=comprehensive, 2=engineering or science concentrated, 3=others
Institution location	Campus location of the institution	Categorical variable: 1=urban, 2=suburban, 3=urban & suburban, 4=small-scale city

985 institution	Whether the institution belongs to Project 985	Dummy variable: 1=yes, 0=no
211 institution	Whether the institution belongs to Project 211	Dummy variable: 1=yes, 0=no
Independent college	Whether the institution is an independent college	Dummy variable: 1=yes, 0=no
Regional Characteristics		
Provincial land area size	Total land area in the province where the institution is located	Continuous variable
Population	Provincial population density	Continuous variable
Unemployment rate	Provincial unemployment rate	Continuous variable
GDP per capita	Gross domestic product per capita in the province	Continuous variable
Consumer Price Index (CPI)	Cost of living index in the province	Continuous variable
Ecological Civilization Index (ECI)	An indicator of environment quality	Continuous variable
Number of higher education institutions	Number of higher education institutions	Continuous variable
Sea coastal province	Whether the province is a sea coastal province	Dummy variable: 1=yes, 0=no

Section 3.4 Quantitative Designs for Estimating Impact of Work Migration on Graduates' Starting Salaries

In order to answer the second research question about impact of work migration on graduates' initial salaries, the Mincer earnings function will be employed. The basic model without selection problem and endogeneity can be written as,

(3.5) $$InW = \beta_0 + \beta_1 X_i + \beta_2 Z_{ij} + \delta_1 M_{ij} + u_{ij}$$

where W is the initial salary for the college graduate, M_{ij} indicates the college graduate's migration behavior, X_i represents a vector of individual characteristics including gender, age, student ability, social economic status, school quality, etc. Z_{ij} includes both personal and location characteristics. Ordinary least-squares (OLS) regression will be estimated first.

However, there might be several sources of bias with regards to the above model (3.5). Though the survey data includes extensive information about college graduates, it is still likely that some important determinants (e.g., unobserved ability) of starting salary are not included in the model. In other words, college graduates' migration behavior is possibly correlated with some unobservable student characteristics such as individual ability. Omission of such variables, particularly when they are also correlated with included explanatory variables will cause bias in identifying the link between starting salary and included variables. Besides omitted variable, another source of bias may come from the endogeneity of migration. According to Roy (1951), migration is endogenous and college graduates self-select into migration. The causality between graduates' starting salary and their work migration behavior could be reverse.

3.4.1 Propensity score matching

Propensity score matching (PSM) provides a means for adjusting for selection bias in observational studies of causal effects. Propensity scores are used in observational studies to reduce selection bias by matching different groups based on the propensity score probabilities, rather than matching individuals on the values of the individual covariates. According to Rosenbaum and Rubin (1983), when many characteristics are used in the matching process, propensity scores can be used to select a comparison group that is similar, on average, to a treatment group along those characteristics. PSM method represents, depending on the particular

method employed either a semi-parametric or non-parametric alternative to linear regression (Smith and Todd, 2004).

Different from OLS models, the PSM strategy does not have to specify the multi-dimensional relationship between explanatory variables and the outcome variable, but it uses a one-number summary of them to control for predictors. Propensity score theory says that rather than controlling for (stratifying on, regressing on, matching on) all the variables in X, it is sufficient to control for just the propensity score, e(x), which is just a one-number summary of X.

This study used propensity scores to select comparison group for treatment group, according to the following steps. The first step of implementing PSM is to determine confounding covariates, which are the ones the researcher cares about balancing across groups. These are the covariates to be considered when checking the overlap and examining the balance.

Then a probit regression will be estimated, where the dependent variable is a binary variable that equals 1 for migrating to work and 0 for not migrating to work, and the predictors are the confounding covariates. The probit regression is used to compute predicted probabilities for each person that they receive the treatment--these are the propensity scores. In this way, a propensity score is assigned to each treatment group member and each potential comparison group member. The propensity score for each individual equals the weighted sum of the individual's values for the characteristics included in the probit model, where the weights are the parameter estimates of the probit model.

Third, for each person in the treatment group, find the person in the control group with the closest propensity score. In this study, three commonly used matching algorithms were employed. The selection process was done with replacement, which means that a potential control group

member could have been selected as many times as she is the best match. Matching with replacement tends to reduce bias relative to matching without placement. The full dataset is reduced to have only the treated observations and only those control observations that are chosen as matches.

In order to make sure that for each treatment group member there is a control group member that is sufficiently similar that can act as an empirical counterfactual, several approaches will be done to investigate the overlap and balance. 1) Plot histograms and check the overlap for the propensity scores in each group separately. 2) Try several different model specifications for estimating propensity scores and compare the balance achieved under each. Some strategies include adding interactions or squared terms, removing variables, or transforming the variables, 3) Test for balance on all covariates initially designated as confounders.

When the matching process is complete, run a multiple regression of outcome on treatment indicator and confounding covariates using weights to force sample to represent matched groups (1 if in treatment group, 0 if not matched). The weights equal the number of times each observation used in the analysis, since the observations in a matched sample are no longer independent.

The most important assumption required for propensity score matching to yield valid causal inferences is ignorability. The ignorability assumption requires treatment assignment is independent of the outcomes (Y) given covariates X. From a practical point of view, ignorability assumption requires observing all covariates X that are simultaneously associated with both treatment status and potential outcomes. Propensity score matching strategy is a more nonparametric way to control for confounding covariates if ignorablity is justifiable. It relies on

weaker assumptions about the way that Y and X are related to each other. However, in empirical study, the ignorability assumption is very strong assumption and we need to assume that we have controlled for all the potential confounders. This strategy does not solve the "omitted variable bias" or "selection bias" problem.

One way to assess how convincing the PSM inferences are is to perform sensitivity analysis, which addresses how much the hidden bias (or the unmeasured covariate) would have to be to alter the conclusions (Rosenbaum, 2002, 2005).

3.4.2 Instrumental variables

Instrumental variable is another identification strategy to be used to address the ignorability assumption problem. To deal with the above mentioned potential bias in the empirical estimation, two instrumental variables for migration will be introduced in the model. An instrumental variable should be an exogenous source of variation, which is correlated with migration but does not affect starting salary through the paths other than migration.

In this study, the following two variables are applied as IVs:

1) The first instrumental variable used in this study is the percentage of graduates migrating to work in each institution. This is a measure of the institution peer effect of work migration. The assumption behind is that if there is a common trend of migrating to work in the institution where a student attends, he/she is more likely to follow the trend, and the probability that he/she would migrate to work after graduation is higher. This variable "percentage of graduates migrating to work in each institution" is partially determines treatment status—migration, but is otherwise unrelated to potential outcomes—starting salary.

2) The rate of employment outflow in the province where the student attended college. The rate of employment outflow in the province where the student attended college is related to migration; meanwhile it doesn't correlate with the unobserved individual ability. Such measure will strongly predict a decision to migrate but does not predict starting salary apart from migration.

Sargan test of overidentifying restrictions will be performed to test the exogeneity of these instrumental variables. Here I_{ij} represents the set of IVs and the first stage regression is to use the instrumental variables I_{ij} to predict the endogenous variable M_{ij} by running regressions of M_{ij} on I_{ij}, which can be written as follows,

(3.6) $$M_{ij}^{*} = \beta_0 + \beta_1 X_i + \beta_2 Z_{ij} + \theta I_{ij} + u_{ij}$$

where $E(\mu_{ij})$ is assumed to be zero, and $Cov(I_{ij}, \mu_{ij})$ is assumed to be zero as well. The predicted values of M_{ij} for each individual is denoted as

(3.7) $$\widehat{M_{ij}} = \widehat{\beta_0} + \widehat{\beta_1} X_i + \widehat{\beta_2} Z_{ij} + \hat{\theta} I_{ij}$$

The second stage is the OLS regression of InW on $\widehat{M_{ij}}$ and other exogenous variables, with corrected standard errors. Then the salary equation would be

(3.8) $$InW = \beta_0 + \beta_1 X_i + \beta_2 Z_{ij} + \delta_1 \widehat{M_{ij}} + u_{ij}$$

The vector of δ_1 will be estimated as the causal effects of work migration on graduates' starting salaries.

Table 3-2 shows the variables that will be included in the wage equations. College graduates' starting salaries are related to students' characteristics, institution attributes, and labor market characteristics. Whether the student has an academic minor is relevant to labor market

performance. It may improve one's competitiveness in the labor market (Du & Yue, 2010). In addition, types of financial aid were also included the models, as they may influence students' incentive of studying and post-college labor market outcomes (Yang, 2011). Besides the variables explained before, the industry dummies for the job and the province dummies where the job is located are also added as covariates in the wage models to control for wage differences between industries and provinces (Titus, 2010). The employer's type is also included in the wage equation. The types include state-owned firms, foreign or co-owned firms, Party or government institutions, or others.

Table 3-2 Definition and Measures of Variables Used in the IV Models

Variable Name	**Definition**	**Core Measurement or Comments**
Dependent Variables:		
Salary	Initial salary indicated in the job offer	Continuous variable; Used in log form
Employment	Initial employment status--whether the student had an offer by the time of survey	Dummy variable: 1=yes, 0=no
Covariates:		
Individual Characteristics		
Work migration	Whether or not a student will migrate to work	Dummy variable: 1=will migrate to work, 0=will not migrate to work
Female	Gender of the student	Dummy variable: 1=female, 0=male
Age	Age of the student in 2011	Continuous variable, calculated by birthday question
Minority	Whether the student is from a minority ethnic group	Dummy variable: 1=minority, 0=Han
From rural area	Whether the student has a rural residence registration (Hukou)	Dummy variable: 1=rural, 0=urban

Only child	Whether the student is the only child in their family	Dummy variable: 1=only child, 0=not only child)
SES score	Constructed index of the socio-economic status of the student's family	Continuous variable; it's a composite score based on parents' years of education, parents' occupations, annual household income, and family wealth measured by assets
Student leader in high school	Whether the student was a leader in high school	Dummy variable: 1=yes, 0=no
Humanities track in high school	Whether the student was on humanities academic track in high school	Dummy variable: 1=yes, 0=no
Arts or sports track in high school	Whether the student was on arts or sports track in high school	Dummy variable: 1=yes, 0=no
NCEE score	National college entrance examination score (rescaled to 1-100), measure of academic ability	Continuous variable
Science or engineering major	Whether the student has a science or engineering major	Dummy variable: 1=yes, 0=no
Economics or management major	Whether the student has an economics or management major	Dummy variable: 1=yes, 0=no
Had a minor	Whether the student had a minor in college	Dummy variable: 1=yes, 0=no
Average course score in college	Student's average score in college	Continuous variable
Preference degree towards one's major	How the student likes his/her major	Ordered categorical variable: 1=not at all, 2= a little bit, 3=somewhat, 4=very much
Passed CET4	Whether the student passed CET4 test (English proficiency measure)	Dummy variable: 1=yes, 0=no
Passed CET6	Whether the student passed CET6 test (English proficiency measure)	Dummy variable: 1=yes, 0=no
Student leader in college	Whether the student was a leader in college	Dummy variable: 1=yes, 0=no
CCP member	Whether the student is CCP member	Dummy variable: 1=yes, 0=no
Has professional certificates	Whether the student has professional certificates	Dummy variable: 1=yes, 0=no
Have worked in college	Whether the student ever worked in college, (including both term time and vacations)	Dummy variable: 1=yes, 0=no
Had merit-based aid	Whether the student had merit-based aid	Dummy variable: 1=yes, 0=no
Had needs-based aid	Whether the student had needs-based aid	Dummy variable: 1=yes, 0=no

Had loan	Whether the student had loan	Dummy variable: 1=yes, 0=no
Institution Attributes		
Institution region	Region of the institution	Categorical variable: 1=municipalities, 2=east area, 3=northeast area, 4=central area, 5=west area
Institution concentration	Concentration/Specialization of the institution	Categorical variable: 1=comprehensive, 2=engineering or science concentrated, 3=others
985 institution	Whether the institution belongs to Project 985	Dummy variable: 1=yes, 0=no
211 institution	Whether the institution belongs to Project 211	Dummy variable: 1=yes, 0=no
Independent college	Whether the institution is an independent college	Dummy variable: 1=yes, 0=no
Labor Market Characteristics		
Industry	The industry in which the student will work after graduation	Categorical variable including transportation, IT, media and sports, finance, manufacturing, etc.
Employment type	Type of the employer	Categorical variable indicating the types of graduate's employer, e.g., state-owned firms, foreign or co-owned firms, Party or government institutions, or others
Work province	The province where the student will work at after graduation	Categorical variable

The instrumental variable strategy could be employed only when the following assumptions are satisfied. The first assumption is exclusion restriction. It says that the IV only affects the outcome through the treatment. In details, there should be no direct effect of IV on outcomes and no treatments associated with IV other than the one of interest.

The second assumption requires the IV has non-zero impact on treatment. An IV isn't useful if it doesn't actually predict the treatment. Specifically, it refers to the situation that is related to a higher probability of college graduates' migration. This assumption is empirically testable by checking the correlation between the IVs and the probability of a college graduate's migration.

The third assumption is monotonicity, which assumes that there were no defiers. The defiers are those who would take the treatment if assigned not to but would not take the treatment if assigned to take it. The fourth assumption indicates that the instrument itself is randomly assigned and the last assumption is stable unit treatment value assumption (SUTVA). It assumes that the treatment status of any unit does not affect the potential outcomes of the other units. And the treatments for all units are comparable (no variation in treatment).

In this study, different tests would be conducted to examine whether the above assumptions are satisfied. Hausman test would also be used to test if the estimates of this IV model are significantly different from the estimates of equation without IV. Heckman method will be employed to adjust for sample selection bias due to student's being offered employment or not.

Section 3.5 Data

This study makes use of a unique dataset obtained from the second part of Chinese College Student Survey-- College Students' Labor Market Data 2011 (CSLM 2011), which was collected by Institute of Education Tsinghua University and China Data Center Tsinghua University in 2011.

As second part of the Chinese College Student Survey, CSLM2011 was expanded on the basis of the first part "NSSE-China," which was a longitudinal research project initiated by the

Graduate School of Education, Tsinghua University in 2009. The "NSSE-China" project includes implementation of Indiana University developed surveys that measure student engagement in China's diverse institutions of higher education. Based on National Survey of Student Engagement (NSSE), launched in 1999 and housed at Indiana University, the NSSE-China instrument asks college students to report their perception of institutional environment, their participation in programs and activities that institutions provide interactions with faculty and other students, time-on-tasks, as well as background characteristics and learning outcomes.

As discussed in first chapter, the Mid- and Long-term Education Reform and Development Plan has established "education quality" as one of the most important educational focuses for the next 20 years in China. The higher education system has formally developed from the external system revolution to the internal quality promotion (China data center, Tsinghua, 2011). With such background, the topic of how to evaluate the quality of higher education and improve higher education effectively has attracted attention from the current Chinese higher educational researchers. This project will continue on an annual basis. Some researchers' findings have been published in 2009 and have had a great impact on both higher education in China and other countries such as America, Germany, Japan, Singapore and Canada. The two basic survey modules in 2010, namely "NSSE-China" and "Follow-up Survey of College Graduates in China", have jointly established the integrated survey and evaluation system on college students' learning and development. It attempts to examine the student's social economic background before they enter college, their learning experience and university life, their behaviors related to employment and future development from different types of universities in different areas. The purpose of starting this survey was to provide all the universities with the data that they need in order to evaluate and improve the quality of education and talent training in a more comprehensive way.

The CSLM2011 conducted multi-stage sampling by using regional variable (three municipalities--Beijing, Shanghai, Tianjin, the northeast, east, central and west regions) and institutions' type variable (first-tier, second tier, associate degree institutions, post-secondary professional schools) as sampling standards. Within each higher education institution, the surveys are distributed randomly based on students' ID. Sampling weights were calculated based on the sampling scheme to ensure national representativeness. For 2011 survey, it covered 49 higher education institutions, 13 are from Beijing/Shanghai/Tianjin, 8 are from the east region, 5 are from northeastern region, 11 are from central region, and 12 are from western region. With regards to the academic ranking of the institutions, there are 8 universities belonging to 985 institutions, 16 are 211 institutions, 23 are second-tier colleges (non-key provincial colleges) and 2 are independent colleges. With regards to academic concentration, there are 15 comprehensive institutions, 21 science and engineering concentrated institutions, 12 institutions concentrated on teacher training and education, agriculture, finance and economics, and political science and law, and 1 university of minority. Science and engineering concentrated institutions are oversampled. Such a sample reflects a reasonable representation of the scope and diversity of higher education in China.

The College Students' Labor Market Data collects information on 8,176 Chinese students who graduated from higher education institutions in 2011. The survey provides comprehensive information on individual characteristics, family background characteristics, high school experience and NCEE performance, college activities, financial situation during college and placement after graduation. The questionnaires were distributed to senior students in late May or June before their graduation. At the time of survey, most students had clear ideas about their

placement after graduation. Therefore, we were able to identify those individuals who decided to accept job offer and migrate to work.

The student sample used in this dissertation study only included students from cohort 2007, i.e. those who entered college in 2007. The purpose of doing so is to eliminate cohort-level differences. There were originally 6,983 students in cohort 2007 and four contracted students were excluded from the sample. These students obtained funding from the military or public schools and they are required to work for their funder after graduation. In other words, they cannot choose their employment on their own like other graduates. Furthermore, one student from Hong Kong and one student who worked in Macau after graduation were excluded, as Hong Kong and Macau are different from the Mainland China in many aspects. The final cohort 2007 sample contains 6, 977 students.

The spatial data that will be employed come from a variety of sources. For example, most economic factors are from China Statistical Yearbook. Educational data are drawn from China Education Statistical Yearbook.

The Center on Chinese Education at Teachers College Columbia University has a collaboration agreement with the Institute of Education at Tsinghua University on a research program on higher education policy in China. Access to the needed survey data for this dissertation study is made possible through this collaboration agreement.

3.5.1 Socio-economic status index (SES) construction

In order to reduce measurement error of self-reported information in student surveys, sometimes multiple questions are asked from different perspectives to evaluate certain variables comprehensively. By using principal components analysis as the extraction methodology, some

variables in the questionnaire are reconstructed to index to avoid collinearity in the empirical regression.

Principal component analysis (PCA) was invented in 1901 by Pearson (1901) and later developed and named by Hotelling in the 1930s. It is a statistical procedure that uses orthogonal transformation to convert a large set of possibly correlated variables into a smaller set of linearly uncorrelated variables called principal components. The transformation is defined in such a way that the first principal component has the largest possible variance in the original variables, and each following component has the highest variance possible under the constraints that it is orthogonal to the preceding components. PCA has been a recently widely used technique to create indices (e.g. Houweling et. al., 2003; Vyas & Kumaranayake, 2006).

We employed PCA to generate the index SES scores and variables used included parents' education level, parents' occupations, parents' work industry, modified annual household income, and type and area of residence dwelling as a measure of household wealth. In the original dataset, most of the above variables except for annual household income and area of residency are categorical variables. As suggested by Vyas & Kumaranayake (2006), categorical variables were not suitable for PCA analysis, because the quantitative scale does not have any meaning. Therefore these variables need recoding to be included in the analysis. Many original variables have more than ten categories. Converting each category to a binary variable will lead to a KMO value far below the "acceptable" threshold value 0.5. Therefore, the categorical variables are then recoded in the following way:

Parents' occupation information was recoded based on combined information of job position, industry, and nature of employer. Three sets of binary variables were created at the

household level respectively. A value of 1 in each variable indicated that at least a parent in the household belonged to that category. The first set of variables described the position or nature of one's occupation. The categories of variable included whether a parent in the household was a manager or leader, a professional staff (i.e. high-skilled worker), an ordinary staff (e.g. office clerk, sales person, etc), self-employed (e.g. small business owner, peddler, etc.), a manual worker or farmer, or unemployed/not in the labor force. The second set of variables described the industry where the parent works. The categories were whether a parent in the household worked in the manufactory industry, retail or service industry, high-income industry including IT and finance, or public service industries including education and medical service. The third set of variables described the nature of the employer. The categories included whether a parent in the household worked for the government, public institutions, enterprises, or for self-owned business. The type of the dwelling was recoded into 6 categories: dwelling in rural area, dwelling in unreconstructed old town community, dwelling in town, dwelling in the residency community of one's employer, ordinary commercial dwelling, and commercial dwelling in high-income community.

Parents' education levels were recoded into a new variable years of schooling based on the following criteria (Du & Yue, 2010): "no school" was recoded as having 0 year of schooling, 'graduated from elementary school' as 6 years of schooling, 'graduated from junior high school' as 9 years of schooling, 'graduated from senior high school or secondary vocational school' as having 12 years of schooling, 'graduated from post-secondary vocational college' as 14.5 years of schooling, and 'college graduate' as 16 years of schooling. As for people who attended graduate schools, 'master degree holders' were coded as having 19 years of schooling, and

'doctoral degree holders' as 22 years of schooling. Such recoding represents the typical length of schooling at each education level in China.

As continuous variables, annual household income and area of dwelling were transformed with natural logarithm to avoid skewness and kurtosis in distribution. Because PCA was sensitive to outliers, the outliers in these variables were deleted. Observations with missing value in these variables were also dropped. After cleaning the data, a total of 5, 231 observations were included in the PCA analysis.

All of the above transformed and recoded variables were included in the PCA analysis. The correlation matrix was investigated and variables that either had too weak (none of the correlation parameters was greater than 0.2) or too strong correlation (any correlation parameter was greater than 0.9) with other variables were dropped. Variables with individual KMO value less than 0.5 were also dropped from the analysis. The decision of which variables to drop was made with an attention to ensure that at least two variables were kept from each of the three sets of variables describing parents' occupation information. There were 14 variables included in the analysis: 4 variables describing parents' occupation position: whether either of the parents is a manager, a professional, an ordinary staff, and a craftsman or farmer; 2 variables describing the nature of employers: whether either of the parents works for the government, and whether either of the parents works for public institutions; and 2 variables describing the industry in which the parents worked: whether either of the parents works in the public service industry, and whether either of the parents works in the service and retail industry; 4 continuous variables: annual household income (in logarithm form), area of dwelling (in logarithm form), father's years of schooling, mother's years of schooling; 2 variables describing type of dwelling: whether the

family live in a rural-dwelling house and whether the family live in an ordinary commercial dwelling unit.

The results of the PCA analysis are included in the Appendix 1. The requirements for conducting valid PCA analysis were satisfied. The null hypothesis of the Bartlett's Test of Sphericity was rejected with a p-value of 0.000. The KMO value of all the variables was 0.805. The determinant of the correlation matrix was 0.019, which is larger than the necessary value of 0.00001. Finally, five principal components with eigenvalues greater than 1 were derived. The first component explained 30.35% of the total variance in the original variables. Therefore, this component was used as the SES score for the individual student.

3.5.2 Missing data

Because of random data collection and non-response issues, some variables have missing values. Missing value is known to some problems like bias, efficiency loss, and incorrect standard errors (Rubin, 1987). There are different missing data mechanisms: missing completely at random (MCAR), missing at random (MAR) and missing not at random (MNAR). MCAR means that whether or not any given value is missing is completely random, or, the probability of missingness is the same for all units (Abayomi, Gelman, & Levy, 2008). This is generally not a plausible assumption. Usually, certain types of people are much more likely than others to have missing data. A weaker condition is MAR, which means missingness depends only on observed values of the variables. In 'Not Missing at Random' cases, both assumptions are violated if the probability of missingness varies and cannot be characterized by the observed values of the items (Rubin, 1976; Little & Rubin, 2002). MCAR and MAR are both ignorable missing data mechanisms. In another word, for these missing data mechanisms we can make inferences using

our data without having to include a model for the missing data mechanism within our analysis model.

When missing values exist in a dataset, available data size shrinks and efficiency decreases; therefore, we always need to deal with missing values in one way or another. As a method to deal with missing data, single imputation is often utilized because it is intuitively attractive. In single imputation, we fill in missing values by some type of predict values,

There are many different strategies to try to resolve the problems associated with missing data: listwise deletion, pairwise deletion, dummy variable adjustment, mean/mode imputation, and multiple imputation, etc. Listwise deletion removes all observations from the dataset that have any missing values. It has been the most often used method to deal with missing data (King et. al., 2001, Ciuk & Pyle, 2009). This method could be employed with MCAR assumption. However, it is impossible to test the assumption of MAR without additional data collection since the information that would be used to make such a test is unavailable in most cases. Also, the reduction of sample size will lead to higher standard errors and may preclude certain types of analyses. For dependent variables in this analysis, listwise deletion method was employed.

"Dummy flag" strategy is widely used by economists and other social scientists for missingness. It creates an indicator for missing value (1 if missing for observation; 0 if observed for observation) and imputes missing values to a constant (such as mean or mode), then includes the indicators for missingness for each variable in regression. In another word, the dummy flag model will use mean/mode imputation (Mean/mode substitution replaces missing value with sample mean or mode and then run analyses as if for all complete cases) and a dummy variable

to flag missingness meanwhile adjusting the variance to compensate for the underestimation of the standard error that typically occurs with unmodified mean imputation.

Though “Dummy flag” method has some disadvantages such as results in biased estimates, it uses all available information about missing observation. When missing values occur for reasons beyond our control, we must make assumptions about the processes that create them and try to resolve the problems it brings. Therefore, “Dummy flag” method was employed for the covariates with missing values before data analysis in this dissertation.

CHAPTER FOUR:
DESCRIPTIVE STATISTICS

In this chapter, descriptive statistics and facts on college graduates' migration categories, graduates' starting salaries information and other individual and institution variables are presented in detail. All the results are first reported without considering the weights, and then represented weighted by sampling weights.

Section 4.1 Descriptive Statistics

When graduating from college, students face several choices: to continue studying in native graduate schools, to search for a job or study abroad etc. The group of graduates who intend to work and search for jobs are the research objects for this study and this group is defined as the "intention to work sample". In selecting the "intention to work sample", there may be a bias introduced into analysis. However, currently this study will only focus on the analysis based on "intention to work sample". Within the intention to work sample, 66.19% graduates (weighted) already had a job offer when taking the survey. Table 4-1 and Table 4-2 present the descriptive statistics of variables that are going to be used in the empirical models. Both of the tables are weighted by sampling weights and both sets of results are derived from variables without replacing missing values. For descriptive statistics, the command used in Stata is "summarize" with ".aweights". Table 4-1 represents the descriptive information for the whole sample (N=6977, the 2007 cohort of college students who graduated in 2011) and Table 4-2 shows the descriptive results for the intention to work sample (N=4984) that will mainly be used in this dissertation.

Table 4-1 Descriptive Statistics of Variables (Weighted), Whole Sample, Year 2011

Variable	Whole Sample (6,977 obs.)		
	Mean/Percentage	Std. Dev.	Missing Rate (%)
Student Characteristics			
Age (years)	22.99	1.00	2.11
Gender (%)		0.50	0.46
Female	47.27		
Male	52.26		
Minority (%) (Yes=1)	5.25	0.22	0.95
In a relationship (%) (Yes=1)	34.47	0.48	4.74
Healthy (%) (No=1)	5.21	0.22	0.90
Risk appetite (%)		0.87	4.43
Risk averse	40.98		
Risk neutral	22.28		
Risk seeking	31.61		
Only child (%) (Yes=1)	36.38	0.48	1.10
Region of residency before college (%)		1.31	2.94
Municipality	8.40		
East	29.17		
Northeast	13.06		
Central	25.81		
West	20.17		
Rural (%) (Yes=1)	43.15	0.50	0.32
Annual household income (in RMB)	46964.20	42248.06	18.26
SES score	-0.15	0.97	22.33
College preference (%)		0.42	5.10
First	72.78		
Others	21.33		
Migrated to college (%) (Yes=1)	28.72	0.46	2.97
National College Entrance Examination (NCEE) score (rescaled to 1–100)	70.41	7.88	12.05
Average score in college	79.64	6.80	22.06
English (%)		0.73	2.90
Not passed CET4	20.24		
Passed CET4	42.48		
Passed CET6	33.37		
Major (%)		1.01	0.21
Liberal arts	14.48		

Social sciences	7.75		
Sciences and engineering	55.47		
Economics and Management	16.73		
Others	5.24		
Whether like major (%)			2.52
Very much	12.31		
Somewhat	47.41		
A little bit	28.38		
Not at all	7.97		
Leader in high school (%) (Yes=1)	41.62	0.49	0.00
Leader in college (%) (Yes=1)	21.78	0.46	0.00
Party member (%) (Yes=1)	29.54	0.46	0.93
Professional certificate (%) (Yes=1)	46.62	0.50	0.00
Ever worked in college (%) (Yes=1)	78.21	0.40	2.26
Had merit aid (%) (Yes=1)	34.13	0.49	0.00
Had needs aid (%) (Yes=1)	21.09	0.47	0.00
Had loan (%) (Yes=1)	27.92	0.45	2.85
Institution Characteristics			
Region of institution (%)		1.36	0.00
Municipality	14.48		
East	25.16		
Northeast	15.53		
Central	25.09		
West	19.73		
Level of institution (%)		0.69	0.00
985 institution (Yes=1)	6.65		
211 but not 985 institution (Yes=1)	12.28		
Non-key institution	69.72		
Independent college	11.44		
Concentration of institution (%)		0.42	0.00
Comprehensive	22.18		
Engineering-concentrated institution	43.34		
Others	34.48		
Location of the campus (%)		1.18	0.00
Urban	30.54		
Urban & suburban	3.89		
Suburban	38.47		
Small city	27.10		
Employment outflow rate (%)	30.38	15.47	0.00
Distance to big city (km)	732.84	456.85	0.00
Sea coastal province dummy (%) (Yes=1)	32.32	0.47	0.00
Consumer Price Index (CPI)	102.67	0.35	0.00

Unemployment rate	3.11	0.81	0.00
Ecological Civilization Index (ECI)	0.73	0.18	0.00
Number of higher education institutions	80.83	23.53	0.00
Provincial GDP per capita ($)	6752.41	2702.82	0.00
Provincial population (10k)	4807.88	2637.76	0.00
Provincial size (10,000 km^2)	16.93	14.91	0.00
Job-Related Characteristics			
Employed (%)	53.20	0.50	0.00
Salary per month (in RMB)	2381.99	1210.58	11.55
Work province unemployment rate (%)	3.20	0.78	11.60
Number of résumés submitted	39.24	279.96	33.42
Distance from institution to workplace	556.81	723.65	11.60
Job industry (%)		4.61	4.42
Agriculture/Fishing/Forestry	2.30		
Mining/Manufactory/Construction	24.55		
Utilities/Energy	5.53		
Transportation/Storage/Postal	3.98		
Telecom/Computer service and software	14.61		
Wholesale/Retail	3.72		
Hospitality/Food services	2.44		
Finance	6.66		
Real Estate	3.68		
Lease & business service	1.94		
Education	7.87		
Medical care	2.70		
Culture/Sport/Social utility	4.38		
Science & research/technology service	5.15		
Water conservancy/Environmental Protection	1.20		
Community service and other services	1.47		
Government/NGO/international organization	1.32		
Other	1.44		
Region of work province (%)			11.60
Municipality	14.72		
East	37.21		
Northeast	6.76		
Central	13.00		
West	14.93		
Number of higher institutions	88.08	24.83	11.60
Provincial GDP per capita ($)	7477.60	2798.26	11.60
Provincial population (10k)	5404.27	3004.39	11.60
Provincial size (10,000 km^2)	17.04	19.96	11.60

According to Table 4-1, the sample weighted average age is 22.99 years. Among the whole sample, 47.27% of students are female, and 52.26% are male. The percentage of the female students in colleges nationwide from 2007 to 2012 was 47.36%, 48.15%, 48.89%, 49.68%, 50.40% and 51.03% respectively (Li & Tian, 2013). The percentage of female students in our sample is consistent with the national figure. In the sample, 5.25% of the students are from a minority group. There are 34.47% of students stated they are in a relationship. 36.38% of the students are an only child. Students with rural registered-residence account for 43.15% of all the students. With regards to the region of residency before college, the percentage of students who lived in the municipality is 8.4%, and is 29.17%, 13.06%, 25.81%, and 20.17% for the east, northeast, central and west areas. The average annual household income (in RMB) is 46964.2 and the average SES score is -0.15. Of the students sampled, 72.78% went to the college that was their first choice and the average National College Entrance Examination (NCEE) score (rescaled to 1--100) for the whole sample is 70.41. In terms of risk, 40.98% of the students belong to the risk aversive type, 22.28% are risk neutral and 31.61% are risk-seeking people.

The percentage of students majoring in liberal arts, social sciences, engineering and sciences, economics and management, and other majors are 14.48%, 7.75%, 55.47%, 16.73%, and 5.24% respectively. The national statistics for the percentage of students majoring in arts, social sciences, engineering and sciences, economics and management, and other majors are 12.09%, 7.62%, 41.61%, 33.82%, and 15.6% respectively. Because the survey was initiated by Tsinghua University, which is an institution with a strong engineering focus, the survey sample included more students with engineering or sciences major. There are 12.31% of students claimed that they liked their major very much, 47.41% somewhat, 28.38% a little bit and 7.97%

not at all. For student achievement in college, the average score in college is 79.64. In the whole sample, 42.48% of the students passed CET4 (National College English Test level 4), 33.37% passed CET6 (National College English Test level 6) and 20.24% did not pass CET4. The percentage of students who are leaders at college is 21.78% and the percentage of students who are Communist Party members is 29.54%. Within the sample, 46.62% of students got at least one professional certificate during college and 78.21% of students have worked either in the college or as intern outside the college. The percentage of the students who were in receipt of merit aid, need aid, loans are 34.13%, 21.09%, and 27.92%, respectively.

From an institution characteristic's perspective, 14.48% of students that study at higher education institutions are located in the municipality, 25.16% in the east, 15.53% in the northeast, 25.09% in the central and 19.73% in the west. Among all the sampled students, 6.65% study at one of the "Project 985" institutions[8], 12.28% of students study at one of the "Project 211" (that are not one of the 985) institutions[9], 69.72% of students study at non-key institutions and 11.44% students study at independent colleges. There are 22.18% of students in comprehensive institutions, 43.34% in engineering-concentrated institutions, and 34.48% in other institutions. The average distance from the higher education institution to the closest big city (Beijing, Shanghai, or Guangzhou) is 732.84km. There are 32.32% of students studying at higher education institutions located in sea coastal provinces. For higher education institution locations, the weighted average provincial GDP per capita is $6752.41, the average population is 48.07 million and the average land area in a province consists of 169,300 km^2.

[8] Project 985 is a constructive project for founding world-class universities in the 21st century conducted by the government of China. In the initial phase, 9 universities were included in the project. During the second phase launched in 2004, the program was expanded to include 39 universities.

[9] Project 211 is the Chinese government's new endeavor aimed at strengthening about one hundred higher education institutions and their key discipline areas as a national priority for the 21st century.

At the time of survey, the average number of résumés that students had submitted was 39 and 53.2% of students sampled had received at least one job offer. According to the survey results, the average starting salary per month (in RMB) reported by the students was 2381.99 Yuan. The average distance from the higher education institution where the student went to college to their future workplace[10] is 556.81km. With regards to work place, 14.72% of the students reported they would work in the municipality, 37.21% in the east, 6.76% in the northeast, 13% in the central and 14.93% in the west areas, respectively.

Table 4-2 Descriptive Statistics of Variables (Weighted), Intention-to-Work Sample, Year 2011

Variable	Intention-to-work sample (4,984 obs.)		
	Mean/Percentage	**Std. Dev.**	**Missing Rate (%)**
Student Characteristics			
Age (years)	23.02	0.99	1.89
Gender (%)		0.50	0.34
Female	45.74		
Male	54.02		
Minority (%) (Yes=1)	5.39	0.23	0.84
In a relationship (%) (Yes=1)	35.02	0.48	4.01
Healthy (%) (Yes=1)	3.89	0.19	0.76
Risk appetite (%)		0.87	3.49
Risk aversive	41.23		
Risk neutral	22.22		
Risk seeking	32.78		
Only child (%) (Yes=1)	34.11	0.48	1.26
Region of residency before college (%)		1.33	2.53
Municipality	9.08		
East	29.97		
Northeast	13.01		
Central	24.15		
West	21.06		

[10] If the student reports the city of the workplace, the distance is calculated as the distance from the city where the student's college locates to the city of the workplace. If the student only reports the province of the workplace, the distance is calculated as the distance from the city where the student's college locates to the capital city of the work province.

Rural (%) (Yes=1)	46.45	0.50	0.30
Annual household income (in RMB)	45662.63	41338.09	17.84
SES score	-0.24	0.94	21.99
College preference (%)		0.43	4.25
First	71.70		
Others	23.70		
Migrated to college (%) (Yes=1)	27.27	0.45	2.55
NCEE score (rescaled to 1–100)	69.82	7.72	11.32
Average score in college	78.62	6.55	22.57
English (%)		0.73	2.73
Not passed CET4	23.26		
Passed CET4	44.72		
Passed CET6	28.53		
Major (%)		1.01	0.12
Liberal arts	13.68		
Social sciences	7.54		
Engineering and sciences	55.44		
Economics and Management	17.71		
Others	5.52		
Whether like major (%)			1.96
Very much	8.62		
Somewhat	47.43		
A little bit	29.60		
Not at all	11.30		
Leader in senior high school (%) (Yes=1)	39.84	0.49	0.00
Leader in College (%) (Yes=1)	20.51	0.46	0.00
Party member (%) (Yes=1)	26.81	0.44	0.98
Professional certificate (%) (Yes=1)	45.12	0.50	0.00
Ever worked in college (%) (Yes=1)	81.00	0.38	1.34
Had merit aid (%) (Yes=1)	30.81	0.48	0.00
Had needs aid (%) (Yes=1)	21.04	0.47	0.00
Had loan (%) (Yes=1)	28.58	0.46	2.01
Institution Characteristics			
Region of institution (%)		1.36	0.00
Municipality	13.28		
East	26.59		
Northeast	15.05		
Central	24.22		
West	20.87		
Level of institution (%)		0.65	0.00
985 institution (Yes=1)	5.15		
211 but not 985 institution (Yes=1)	10.82		

Non-key institution	72.79		
Independent college	11.24		
Concentration of institution (%)		0.41	0.00
Comprehensive	21.16		
Engineering-concentrated institution	44.10		
Others	34.74		
Location of the campus (%)		1.18	0.00
Urban	33.39		
Urban & suburban	3.08		
Suburban	39.23		
Small city	24.30		
Employment outflow rate (%)	30.09	15.73	0.00
Sea coastal province dummy (%) (Yes=1)	33.93	0.47	0.00
Consumer Price Index (CPI)	102.67	0.34	0.00
Unemployment rate	3.11	0.79	0.00
Ecological Civilization Index (ECI)	0.73	0.18	0.00
Number of higher education institutions	82.21	21.86	0.00
Distance to big city (km)	737.85	453.52	0.00
Provincial GDP per capita ($)	6701.09	2627.26	0.00
Provincial population (10k)	4727.16	2564.08	0.00
Provincial size (10,000 km^2)	16.82	14.84	0.00
Job-related Characteristics			
Employed (%)	66.19	0.47	0.00
Salary per month (in RMB)	2376.94	1207.63	10.54
Work province unemployment rate (%)	3.20	0.78	10.37
Number of résumés submitted	17.48	23.73	17.13
Distance from institution to workplace	557.88	724.48	10.37
Industry (%)		4.60	3.89
Agriculture/Fishing/Forestry	2.19		
Mining/Manufactory/Construction	24.76		
Utilities/Energy	5.37		
Transportation/Storage/Postal	4.04		
Telecom/Computer service and software	14.71		
Wholesale/Retail	3.70		
Hospitality/Food services	2.22		
Finance	6.86		
Real Estate	3.87		
Lease & business service	2.00		
Education	8.10		
Medical care	2.73		
Culture/Sport/Social utility	4.37		
Science & research/Technology service	5.20		

Water conservancy/Environmental Protection	1.13		
Community service and other services	1.42		
Government/NGO/International organization	1.19		
Other	1.48		
Region of work province (%)			10.37
Municipality	14.66		
East	37.54		
Northeast	7.11		
Central	13.23		
West	15.37		
Number of higher institutions	87.79	24.62	10.37
Provincial GDP per capita ($)	7449.64	2789.85	10.37
Provincial population (10k)	5371.94	2982.76	10.37
Provincial size (10,000 km^2)	17.15	20.00	10.37

In order to examine the second research question on the impact of work migration on post-college labor market outcomes, this study only uses a subgroup of college graduates—a group who has an intention to work after graduation (intention-to-work sample) and already has job offers with observable salary values. In the whole sample, about 5% of graduates plan to study abroad and 21% apply for graduate school. Because they do not intend to search for jobs, these students were excluded from the 'intention-to-work' sample. There is another 7% of the whole sample who claimed that they did not have a clear plan at the time of survey. These students were included in the sample if they took actions to look for jobs. The final 'intention-to-work' sample included 4,984 students. Table 4-2 shows the descriptive results for the 'intention-to-work' sample, the sample weighted average age is 23.02. Among the sample, 45.74% of students are female, and 54.02% are male. In the 'intention-to-work' sample, 5.39% of students are from a minority group. Over thirty-five percent of students (35.02%) claimed they were in a relationship. Students reporting they were not healthy amounted to 3.89%. Students with rural registered-residence accounted for 46.45% of all students within the 'intention-to-work' sample.

With regards to the region of residency before college, the percentage of students who lived in the municipality is 9.08%, and 29.97%, 13.01%, 24.15%, and 21.06% for east, northeast, central and west areas. The average annual household income (in RMB) is 45662.63 and the average SES score is -0.24. Within the intention-to-work sample, 71.70% of the students went to the college that was their first choice and the average NCEE score (rescaled to 1--100) is 69.82. There are 41.23% of students that belong to the risk aversive type, and 22.22% are risk neutral and 32.78% categorized as risk-seeking people.

The percentage of students majoring in liberal arts, social sciences, engineering and sciences, economics and management, and other majors are 13.68%, 7.54%, 55.44%, 17.71%, and 5.52%. There are 8.62% of students who claimed they liked their major very much, 47.43% somewhat, 29.60% a little bit and 11.30% not at all. For student achievement in college, the average score is 78.62. There are 44.72% of students who passed CET4, 28.53% passed CET6 and 23.26% did not pass CET4. The percentage of students who are leaders at college is 20.51% and the percentage of students who are CCP members is 26.81%. At least one professional certificate was obtained by 45.12% of the students during college and 81% of students worked in college. The percentage of the students who received merit aid, need aid, or loans is 30.81%, 21.04%, and 28.58%, respectively.

Over thirteen percent of the students (13.28%) within the sample that study at higher education institutions are located in the municipality, 26.59% in the east, 15.05% in the northeast, 24.22% in the central and 20.87% in the west areas. Among all the students in the ‘intention-to-work’ sample, 5.15% study at one of the Project 985 institutions, 10.82% of the students study at one of the Project 211 (that are not one of the 985) institutions, 72.79% of the students study at non-key institutions and 11.24% of students study at independent colleges. There are 21.16% of

the students studying in comprehensive institutions, 44.10% in engineering-concentrated institutions, and 34.74% in other institutions. The average distance from the higher education institution to the closest big city (Beijing, Shanghai, or Guangzhou) is 737.85km. For where the higher institution is located, the weighted average provincial GDP per capita is $6701.09, the average population is 47.27 million and the average land area is 168,200 km^2.

At the time of the survey, the average number of résumés that students submitted was 17.48 and 66.16% of the students sampled had received at least one job offer. According to the survey results, the average starting salary (in RMB) per month is 2376.94 Yuan. The average distance from the higher institution where the student goes to college to the future workplace is 557.88km. With regards to the work place, 14.66% of the students will work in municipality, 37.54% in the east, 7.11% in the northeast, 13.23% in the central and 15.37% in the west areas, respectively.

As shown in Table 4-2, the overall missing rate of variables in this "Intension-to-work" sample is not high and missing values are therefore not a big problem for this analysis. The missing rates of most covariates are below 5%. Three covariates, "NCEE score", "annual household income", and "number of résumés submitted" have missing rates between 10% and 20%. The two variables, "SES score" and "average score in college" have missing rates around 22%. In terms of the dependent variables, the missing rates for "migration" and "salary per month" are 10.37% and 10.54% respectively. This rate is calculated for those who had obtained employment offer by the time of survey in the "intention-to-work" sample. Observations with missing values among the dependent variables were deleted from the analysis. As discussed in a previous chapter, missing values in covariates are treated with the "dummy flag" method.

Section 4.2 Work Migration Results

As shown in the following descriptive table (weighted), 27.51% of students who had offer(s) at the survey time reported that they would migrate to work after graduation (sum of return migrants and repeat migrants) while 36.29% of students would be working in the province where his/her higher education institution is located.

Table 4-3 Descriptive Statistics of Migration to Work (Weighted)

Migration to Work	Frequency	Percentage
College stayers	1809	36.29
Return migrants	267	5.36
Repeat migrants	1104	22.15
Missing	368	10.37
Do not know	1436	25.83
Total	**4984**	**100**

When considering the flow direction of those 27.51% college graduates, more detailed categories are used to capture such information. Over five percent of the college graduates would go back to work in their hometown (5.36%), while 22.15% of college graduates chose to work in a province that is neither their hometown nor where their college is based. At the time of the survey, 25.83% students did not have any job offers; therefore, we do not know their migration decisions.

Table 4-4 presents the work migration behavior among different groups. After graduation 27.51% of the students would migrate to work. Among all the students within the sample, 41.5%

of the male students would migrate to another province for work while only 15.8% of the female students would migrate to work. Within the only child group, 21.55% of students will migrate to work, while 30.98% of students, who are not only children, will migrate to work. For students who migrated for college study, 49.41% will migrate to work. For students who went to the college in the same province as they live, 19.63% of them will migrate to work.

In terms of the ranking of the higher education institutions, for Project 985 institutions, Project 211 but not 985 institutions, non-key institutions and independent institutions, the percentage of the students that will migrate to work are 49.81%, 32.65%, 24.59%, and 31.25%, respectively. From a different program concentration/focus perspective, for comprehensive institutions, engineering-concentrated institutions and institutions with other concentrations, the percentage of students that will migrate to work are 17.65%, 45.22%, and 11.03% respectively.

The percentages of students who will migrate to work are also different in terms of location of higher institution. Central China has the highest percentage, with 42.67% of the students finding a job outside the province where their higher institution is. The percentages are 9.37%, 10.19%, 33.20% and 39.42% for the municipality, east, northeast, and west respectively. The higher migration percentage for students in central and west areas is consistent with the fact that there are fewer job opportunities for college graduates in these places. Students with different major have various migration behavior, the percentages of the students that migrate to work are 19.14%, 8.08%, 35.89%, 20.68%, and 15.23% for liberal arts students, social sciences students, engineering and sciences students, economics and management students, and students with other majors, respectively.

Table 4-4 Incidence of Work Migration in China, Year 2011

(Sample size = 4984, weighted)

	Number of Students Migrating to Work	Total Number	Migrate to Work (percentage)
Overall	1371	4984	27.51%
By gender			
Female	425	2690	15.80%
Male	945	2277	41.50%
By Only child status			
Only child	369	1712	21.55%
Not only child	994	3209	30.98%
By previous migration behavior			
Migrate to college	673	1362	49.41%
Stay at college	686	3495	19.63%
By health status			
Healthy	1314	4752	27.65%
Not healthy	47	194	24.23%
By ranking level of institution			
"Project 985" institution	128	257	49.81%
"Project 211" but not "Project 985" institution	176	539	32.65%
Non-key institution	892	3628	24.59%
Independent institution	175	560	31.25%
By concentration of institution			
Comprehensive institution	186	1054	17.65%
Engineering-concentrated institution	994	2198	45.22%
Institution with other concentration	191	1732	11.03%
By region of institution			
Municipality	62	662	9.37%
East	135	1325	10.19%
Northeast	249	750	33.20%
Central	515	1207	42.67%
West	410	1040	39.42%
By student major			
Liberal Arts	125	653	19.14%
Social Sciences	32	396	8.08%

Engineering and sciences	984	2742	35.89%
Economics and management	183	885	20.68%
Others	46	302	15.23%

Section 4.3 Monthly Starting Salary

According to Table 4-5, the average starting salary for all subjects in the "intention-to-work" sample is RMB 2376.94. The average starting salary for college stayers is RMB 2177.12, the average starting salary for return migrants is RMB 2464.21 and the average starting salary for repeat migrants is RMB 2695.06, respectively. It shows that the average monthly salary for return migrants and repeat migrants are higher than college stayers, respectively.

Table 4-5 Average Starting Salary by Groups (weighted)

Types of Migration	Mean Monthly Starting Salary (Yuan)
College stayers	2177.12
Return migrants	2464.21
Repeat migrants	2695.06
All subjects	2376.94

Section 4.4 Correlations between Covariates

To detect potential multicollinearity, the correlations among explanatory covariates were checked and the correlation matrix results are shown in Table 4-6. The table presents the pair-wise Pearson correlation coefficients between the explanatory covariates. The correlation between outcome variables and explanatory covariates is not presented here. From the

correlation results shown below, most of the coefficients between explanatory covariates are below 0.3, indicating that there are no strong correlations between these variables. However, some correlation coefficients are above 0.5. Specifically, the correlation coefficient between SES score and "rural" is -0.62. It is explainable because "whether the student is from a rural area" is highly correlated with the rural dwelling variable which was used to construct the SES index. The Pearson correlation coefficient between "major in engineering or sciences in college" and "humanities track in high school" is -0.60. This is because students were divided into three tracks—humanities track, science track, and arts or sports track during high school. Rarely would students on humanities track choose science or engineering majors at college. Although the correlation coefficients are higher than 0.5, when these variables are included in regression models, the variation inflation factors (VIF) of these variables are all smaller than 5. Therefore, multicollinearity is not likely to be a problem here.

Another high correlation coefficient (0.73) was found between the two variables "percentage of work migration students within one's institution" and "institutions with engineering concentration". This may be because students from institutions with a concentration of engineering programs have potentially more employment opportunities and are therefore more likely to find jobs around the country. This increases the probability of the students' work migration within one's institution. Also, the Pearson correlation coefficient between "percentage of work migration students within one's institution" and "whether institution is located in the central or west areas" is 0.58. This is because there are more job opportunities in East China and students are therefore more likely to move to cities in East China for jobs. Though the Pearson correlation coefficients between these variables are relatively high, the VIF (variance inflation

factor) of each explanatory variable in the regression analysis is below 5. Therefore, multicollinearity is not an issue here.

Table 4-6 Correlation Coefficients of Variables (weighted)

	Migrated to College	Unhealthy	In a Relationship	Risk Appetite	Age	Female	Minority
Migrated to college	1						
Unhealthy	-0.0380**	1.000					
In a relationship	-0.014	0.0584**	1.000				
Risk appetite	-0.015	-0.0311**	0.000	1.000			
Age	-0.019	0.0427**	0.0680**	0.019	1.000		
Female	-0.0706**	0.004	0.0373**	-0.0909**	-0.0449**	1.000	
Minority	0.0790**	-0.018	0.005	-0.003	0.016	0.018	1.000
Have worked in college	-0.0249*	0.021	0.0292*	0.003	0.0705**	0.1282**	-0.0344**
Only child	0.0358**	-0.0437**	0.007	-0.001	-0.1296**	0.018	0.0682**
Rural	-0.0264*	-0.0418**	-0.002	-0.007	0.1307**	-0.0978**	-0.0521**
SES score	0.0770**	-0.0380**	0.0322**	-0.015	-0.1978**	0.1041**	0.0700**
Major in engineering or sciences	0.0537**	0.008	-0.018	0.0399**	-0.003	-0.3894**	-0.0250*
Major in economics or management	-0.004	-0.0575**	0.022	0.009	-0.0365**	0.0942**	0.004
Leader at senior high school	0.0351**	0.0898**	0.0543**	0.0293*	0.0355**	0.000	0.019
Humanities track in high school	-0.0886**	0.0358**	-0.005	-0.0304*	-0.004	0.3634**	-0.010

Arts or sports track in high school	0.0311**	-0.006	0.023	0.0267*	0.0368**	0.0382**	0.003
NCEE score	0.1151**	-0.018	-0.008	-0.0509**	-0.0972**	-0.0363**	-0.0861**
Average score in college	-0.004	0.0469**	0.020	-0.0311**	0.0355**	0.2446**	-0.003
Has a minor	0.013	0.001	-0.002	0.0368**	0.007	0.0427**	0.0262*
Preference degree towards one's major	0.020	-0.0566**	0.0491**	0.012	0.0264*	0.0250*	0.0398**
Passed CET6	0.0653**	0.0503**	0.0491**	-0.0700**	-0.1086**	0.2035**	-0.0679**
Passed CET4	-0.0393**	-0.008	-0.015	0.0313**	0.0507**	-0.0735**	0.006
Student leader in college	-0.012	0.0657**	0.1240**	-0.0238*	0.004	0.0593**	-0.004
CCP member	0.0391**	0.0722**	0.0767**	-0.015	0.0323**	0.1068**	-0.0514**
Have certificates	-0.0291*	0.0243*	-0.009	-0.0301*	0.005	0.013	-0.011
Had merit aid	0.0499**	0.0712**	0.0708**	0.002	-0.002	0.1750**	-0.006
Had need aid	0.013	0.0545**	0.0592**	-0.0344**	0.1350**	0.0446**	-0.019
Had loan	-0.023	0.1034**	0.009	0.019	0.1122**	-0.0708**	-0.0253*
Comprehensive institution	0.009	-0.0355**	0.0240*	-0.0273*	-0.0680**	0.015	0.018
Institution with engineering concentration	0.1839**	-0.0889**	-0.0272*	0.0417**	-0.0426**	-0.2778**	0.014
985 institution	0.2458**	-0.021	0.0244*	-0.0260*	-0.0692**	-0.021	0.0496**
211 institution	0.1501**	-0.0282*	0.000	-0.0253*	-0.0743**	-0.0354**	0.0291*
Institution is located in central or west areas	0.0353**	0.0511**	0.014	0.0260*	0.0331**	-0.1998**	-0.0600**

Independent institution	-0.010	-0.016	-0.0714**	0.0489**	0.0438**	-0.1088**	0.0302*
Percentage of work migration students within one's institution	0.2185**	-0.1023**	-0.004	0.018	-0.0325**	-0.2496**	0.012
Employment outflow rate	0.1483**	0.020	-0.005	0.0351**	0.022	-0.1462**	0.0519**

*P < 0.05 **p<0.01

Table 4-6 Correlation Coefficients of Variables (weighted)--Continued

	Have Worked in College	Only Child	Rural	SES Score	Major in Engineering or Sciences	Major in Economics or Management	Leader at Senior High School
Have worked in college	1						
Only child	-0.1688**	1					
Rural	0.1342**	-0.4579**	1				
SES score	-0.1128**	0.4878**	-0.6192**	1			
Major in science or engineering	-0.0770**	-0.0565**	0.1096**	-0.0905**	1		
Major in economics or management	0.0117	0.0643**	-0.0499**	0.0753**	-0.4952**	1	
Leader at senior high school	0.0392**	0.0619**	-0.0551**	0.0343**	-0.0467**	0.0271*	1
Humanities track in high school	0.0822**	-0.0241*	-0.0654**	0.0576**	-0.6006**	0.2165**	0.0288*
Arts or sports track in high school	-0.0083	0.0894**	-0.0771**	0.0525**	-0.2079**	-0.0951**	0.022
NCEE score	-0.0416**	-0.0782**	0.0963**	-0.0431**	0.1496**	0.0313**	0.0148
Average score in college	0.0192	0.0206	-0.0280*	0.0053	-0.1716**	-0.0015	0.1359**

Has a minor	0.0049	0.0786**	-0.0803**	0.1030**	-0.0541**	0.0218	0.0439**
Preference degree towards one's major	0.0229	0.0490**	0.0269*	0.0418**	-0.0598**	0.0064	0.0782**
Passed CET6	0.0306*	0.0037	-0.0504**	0.0802**	-0.0734**	0.0603**	0.0198
Passed CET4	-0.0227	0.0037	0.0355**	-0.0437**	0.0813**	-0.0246*	0.011
Student leader in college	0.0540**	0.0473**	-0.0892**	0.0885**	-0.0620**	0.0203	0.1737**
CCP member	0.0455**	-0.0819**	0.0576**	-0.0424**	-0.0241*	-0.0117	0.1563**
Have certificates	0.0258*	-0.015	0.0209	-0.0283*	-0.0139	0.0397**	0.0201
Had merit aid	0.1275**	-0.0882**	0.0493**	-0.0323**	0.0323**	-0.0167	0.1335**
Had need aid	0.1544**	-0.2271**	0.2198**	-0.2649**	0.0047	-0.0625**	0.0417**
Had loan	0.1442**	-0.2714**	0.2532**	-0.2908**	0.0665**	-0.0668**	0.0475**
Comprehensive institution	0.0296*	0.0653**	-0.0834**	0.1147**	-0.0187	0.0872**	-0.0057
Institution with engineering concentration	-0.1527**	-0.019	0.0643**	-0.0270*	0.3066**	-0.0001	0.0023
985 institution	-0.0255*	0.0633**	-0.0308**	0.0913**	0.0304*	-0.0034	0.0404**
211 institution	-0.0365**	0.0360**	-0.0262*	0.0713**	0.0485**	0.0257*	0.0091
Institution is located in central or west areas	-0.0375**	-0.1850**	0.1478**	-0.1290**	0.1049**	0.0189	0
Independent institution	-0.1081**	0.0583**	-0.0705**	0.0128	-0.0207	-0.0125	0.0086
Percentage of work migration students within one's institution	-0.1258**	-0.0732**	0.1181**	-0.0829**	0.2477**	0.0053	-0.0064
Employment outflow rate	-0.1026**	0.0780**	-0.0685**	0.0780**	0.0949**	0.0288*	0.0270*

*P < 0.05 **p<0.01

Table 4-6 Correlation Coefficients of Variables (weighted)--Continued

	Humanities Track in High School	Arts or Sports Track in High School	NCEE Score	Average Score in College	Has a Minor	Preference Degree towards One's Major	Passed CET6
Humanities track in high school	1						
Arts or sports track in high school	-0.1425**	1					
NCEE score	0.0452**	-0.4767**	1				
Average score in college	0.1378**	0.0849**	0.0041	1			
Has a minor	0.0336**	0.0263**	0.0177	0.0694**	1		
Preference degree towards one's major	0.0201	0.0656**	-0.0319**	0.1622**	0.0028	1	
Passed CET6	0.1312**	-0.1519**	0.3512**	0.1798**	0.0386**	0.0428**	1
Passed CET4	-0.0583**	-0.0441**	-0.0566**	-0.0500**	-0.0079	-0.0285*	0.4081**
Student leader in college	0.0919**	-0.0188	0.0515**	0.1418**	0.0611**	0.0991**	0.0894**
CCP member	0.0474**	0.0069	0.1017**	0.2060**	0.019	0.0925**	0.1453**
Have certificates	-0.0002	-0.0047	0.0066	0.0389**	0.0097	0.0055	0.0453**
Had merit aid	0.01	-0.0316**	0.0555**	0.3331**	-0.003	0.1269**	0.1892**
Had needs aid	0.0296*	-0.0027	0.0038	0.1075**	-0.0657**	0.0798**	0.0557**
Had loan	-0.0326**	-0.0193	-0.0045	0.0097	-0.0095	0.0371**	-0.0304*

Comprehensive institution	0.0002	-0.0057	0.1533**	0.0354**	0.0146	0.0422**	0.1144**
Institution with engineering concentration	-0.2871**	-0.0705**	0.0892**	-0.1414**	0.0093	-0.0390**	-0.0625**
985 institution	-0.0494**	-0.0205	0.2799**	0.0594**	-0.0103	0.0271*	0.1506**
211 institution	-0.0600**	-0.0202	0.2915**	0.0049	0.0332**	0.0155	0.1443**
Institution is located in central or west areas	-0.0847**	0.0199	-0.0284*	-0.0633**	-0.0368**	-0.0228	-0.0434**
Independent institution	0.0167	0.0412**	-0.3491**	0.0657**	-0.008	-0.0215	-0.1858**
Percentage of work migration students within one's institution	-0.2378**	-0.0487**	0.1102**	-0.1501**	0.0054	-0.0336**	-0.0250*
Employment outflow rate	-0.1412**	0.0492**	-0.2719**	0.0215	-0.0091	-0.0012	-0.1741**

*P < 0.05 **p<0.01

Table 4-6 Correlation Coefficients of Variables (weighted)--Continued

	Passed CET4	Student Leader in College	CCP Member	Have Certificates	Had Merit Aid	Had Needs Aid	Had Loan
Passed CET4	1	0					
Student leader in college	-0.0283*	1					
CCP member	-0.0311**	0.2176**	1				
Have certificates	0.0354**	0.0372**	0.0449**	1			
Had merit aid	-0.0305*	0.1572**	0.3475**	0.0571**	1		
Had need aid	-0.0279*	0.0326**	0.1331**	0.0296*	0.1974**	1	

Had loan	0.0196	0.013	0.0583**	0.0184	0.0574**	0.2857**	1
Comprehensive institution	-0.0617**	0.0407**	-0.0443**	-0.0710**	-0.0147	-0.0484**	-0.1008**
Institution with engineering concentration	0.0400**	-0.0548**	-0.0284*	0.0105	-0.0329**	-0.0451**	0.0511**
985 institution	-0.0663**	0.0278*	0.0471**	-0.0366**	0.0375**	0.0356**	-0.0315**
211 institution	-0.0527**	0.0129	0.0704**	-0.0214	0.0198	0.0111	-0.0267*
Institution is located in central or west areas	-0.003	-0.0157	0.1900**	0.0253*	0.0145	0.1107**	0.1941**
Independent institution	0.0667**	-0.0381**	-0.1299**	0.0637**	-0.0337**	-0.0272*	-0.0288*
Percentage of work migration students within one's institution	0.0250*	-0.0664**	0.0242*	0.0373**	-0.0266*	0.0137	0.0914**
Employment outflow rate	0.0425*	-0.0515*	-0.0056	0.0161	-0.0011	0.0274*	0.0775**

*P < 0.05 **p<0.01

Table 4-6 Correlation Coefficients of Variables (weighted)--Continued

	Comprehensive Institution	Institution with Engineering Concentration	985 Institution	211 Institution	Institution is Located in Central or West areas	Independent Institution	Percentage of Work Migration Students Within One's Institution	Employment Outflow Rate
Comprehensive institution	1							
Institution with engineering concentration	-0.4669**	1						
985 institution	0.2958**	-0.1419**	1					

211 institution	0.0973**	0.0868**	-0.0991**	1				
Institution is located in central or west areas	-0.1628**	0.3614**	0.0091	0.0285*	1			
Independent institution	-0.1919**	0.1455**	-0.0952**	-0.1345**	0.1342**	1		
Percentage of work migration students within one's institution	-0.2500**	0.7318**	0.1547**	0.0326**	0.5784**	0.0584**	1	
Employment outflow rate	-0.2787**	0.3909**	0.0033	-0.0531**	0.4888**	0.3990**	0.4692**	1

*P < 0.05 **p<0.01

CHAPTER FIVE:
DETERMINANTS OF WORK MIGRATION

This chapter examines the first research question regarding the determinants of work migration. To do so, the probit model and the multinomial logit model are estimated to identify significant factors that influence the choice of work migration. The weighted intention-to-work student sample is used in the estimation of the two models. For simplicity, the coefficients of missing value dummy variables will not be presented. The results for the two models are given respectively in Section 5.1 and Section 5.2. Section 5.3 gives a summary of the findings.

Section 5.1 Probit Model

In this section we present results for the probit model of college graduates' work migration choice, between "to migrate" and "not to migrate". The model is formally described by equation (3.1) in Chapter 3. Table 5-1 presents the marginal effects of the exogenous regressors on work migration and all standard errors shown are robust. There are three columns in the table, representing the results for three different equations. Equation (1) represents the results for probit regression of work migration on covariates of individual characteristics and institutional characteristics only. Compared with Equation (1), Equation (2) includes additional economic and labor-market variables for the province where the institution is located. And compared with Equation (2), Equation (3) includes additional non-economic variables for the province where the institution is located.

Equation (1) shows that previous migration behavior has a statistically significant positive influence on work migration. If one migrated to attend college, he/she will be more likely to migrate to work after graduation. Being female is statistically significantly associated with lower probability of work migration. Compared with female students, male students are more likely to move to another province for work. If the student is a student leader at high school, then the student is statistically significantly more likely to migrate to work. As for the student's academic achievement, neither the coefficients on student's average score in college or the NCEE score are statistically significant. As for English proficiency, both the coefficients on passing CET4 and passing CET6 are statistically significantly positive, which means students who passed CET4 and CET6 are more likely to migrate to work compared with those who did not pass the CET4. Holding professional certificates exerts a significantly positive impact on graduate's work migration.

Among college experience variables, a student's major significantly influence their migration status. Compared with students with a humanities major, graduates with a science or engineering major will have a higher probability of migrating to work, all things being equal. Also, students on an art and sports track in high school are more likely to migrate to work. In terms of institutional characteristics, both academic ranking and institution concentration have a significant impact on graduates' work migration. Students from Project 985 institutions have a higher probability of migrating to work than those in non-key institutions. Compared with students in institutions with other concentrations, students from engineering-concentrated institutions are significantly more likely to migrate to work. Compared with institutions located in other regions, graduates from institutions in central and west areas have a higher probability of migrating to work.

Other variables including age, minority, health, in a relationship, risk appetite, only child, from rural area, SES score, economics or management major, humanities academic track in high school, NCEE score, average score in college, student leader in college, the degree of preference towards one's major, whether student is a CCP member, comprehensive institution, and independent college are not significant predictors.

Previous literature indicates that migration has been found to respond to relative labor market opportunities and local economic conditions for the working-age population. Therefore some economic and labor market variables for the province where the institution is located were added to the probit model. Specifically, the variables of provincial GDP, population, area size, unemployment rate, and Consumer Price Index (CPI) were included in Equation (2) in Table 5-1. The results for Equation (2) are similar to those for Equation (1).

Consider the results in Equation (2). In terms of individual characteristics, study migration behavior has a statistically significant positive influence on work migration. Previous study migration behavior is associated with an increase in the probability of work migration. Being female is statistically significantly associated with lower probability of work migration. If the student is a student leader at high school, then the student is statistically significantly more likely to migrate to work.

Among college experience variables, a student's major significantly influences their migration status. Compared with students with humanities major, graduates with a science or engineering major are more likely to migrate to work, all things being equal. As for the English proficiency, both the coefficients on passing CET4 and passing CET6 are statistically significantly positive. Compared with students who did not pass the CET4 and CET6, students

who did pass the CET4 and CET6 are more likely to migrate to work. In terms of institution characteristics, both academic ranking and institution concentration have a significant impact on graduates' work migration. Students from 985 institutions or 211 institutions have a higher probability of migrating to work than those in non-key institutions. Compared with students in institutions with other concentrations, students from engineering-concentrated institutions are significantly more likely to migrate to work. After adding the economic and labor market variables (unemployment rate, CPI, GDP, population, province area size), the variable "institution is located in central and west areas" is no longer a significant predictor of work migration.

For economic and labor market information, provincial GDP, population, and area size all have a statistically significant negative impact on work migration, which means graduates from provinces with a higher GDP, larger population and larger area size will be less likely to migrate to another province to work. Higher GDP is an indicator of positive economic conditions in the initial location (where the institution is located). A larger provincial area size increases the potential for moving costs and difficulty. A larger population means more job opportunities, at least in service industries. These factors all have a negative influence on work migration.

Some non-economic factors also play a role in determining new graduates' work migration. Research has found that amenities associated with climate and other factors have had an impact on the direction of moves (Kodrzycki, 2001). In addition, cultural atmosphere may also have an impact on migration. Therefore, in the third column of Table 5-1, three additional variables were added to the probit model to examine the determinants of work migration. These variables include dummy for sea coastal province, ECI (Ecological Civilization Index), and number of higher education institutions. The sea coastal province dummy and ECI are indicators of local

climate conditions and number of higher education institutions is an indicator for the cultural atmosphere.

The Ecological Civilization Index is a measure of the extent that the natural environment of an area is contaminated, and its impact on human health status. As a new mode of civilization that reflects a level of harmony between humans and nature, ecological civilization represents a major conceptual advance for the development of human civilization. The higher the ECI, the more developed human and nature civilization is. In detail, the ECI consists of two parts: the Ecological Efficiency Index (Eco-efficiency index or EEI) and the Environmental Quality Index (EQI). The EEI and EQI were weighted and then calculated to become the 2013 revised ECI (Liu, 2014). The EEI measures the degree and efficiency of the region's ecological resources consumption to achieve economic development in the region. Gross domestic product (GDP) being equal, the smaller the impact of economic development on the natural environment, the higher the EEI is. The EQI characterizes the quality of the living environment from its air quality point of view and is directly related to the quality of people's lives. To calculate the EQI, the air quality index (Air Quality Integrated Index, AQII) was adopted as the core indicator of the environmental quality of a region.

Consider the results in Equation (3). Similar to the regression results in the first and second columns, previous migration behavior has a statistically significant positive impact on work migration. Having migrated before is associated with an increase of about 37% in the probability of work migration.

As for students' individual characteristics, being female is statistically significantly associated with a 5.7% lower probability of work migration. If the student is a student leader at

high school, which is an indicator of ability, then the student is statistically significantly 7.5% more likely to migrate to work. Among college experience variables, a student's major again significantly influences their work migration status. Compared with students with a humanities major, students with a science or engineering major will have an 11% higher probability of migrating to work, all things being equal. As for English proficiency, both the coefficients on passing CET4 and passing CET6 are statistically significantly positive, which means compared with those who did not pass the CET test, students who passed CET4 and CET6 are associated with a 10% to 14% higher probability of migrating to work.

In terms of institution characteristics, both academic ranking and institution concentration have significant impact on graduates' work migration. Students from 985 or 211 institutions have a 28.3% and 8.8% higher probability of migrating to work than those in non-key institutions, respectively. Compared with students in institutions with other concentrations, students from engineering-concentrated institutions are associated with a 23% higher probability of migrating to work. After adding the provincial economic and labor market variables (unemployment rate, CPI, GDP, population, provincial area size), the institution location variable is not significant any more.

For provincial economic and labor market conditions, provincial GDP, population, and area size all have a statistically significant negative impact on work migration, which means graduates from provinces with higher GDP, larger population and greater area size will be less likely to migrate to another province to work.

In terms of non-economic factors that were added to the third equation, the ECI has a statistically significant negative effect, which means that the higher the ECI in the region, the

lower the possibility that the student will move to another province to work. Specifically, a one point increase in ECI is associated with an increase of 31.8% in the probability of work migration.

Besides the insignificance of variables discussed before (age, minority, health, in a relationship, risk appetite, only child, from rural area, SES score, economics or management major, humanities track in high school, art & sports track in high school, NCEE score, average course score in college, preference degree towards one's major, student leader, CCP member, has professional certificates, institution located in Central or West China, comprehensive institution, independent college, unemployment rate, CPI), the sea coastal province dummy and the number of higher education institutions were not statistically significant in the regression model.

Table 5-1 Probit Model on Work Migration
Dependent Variable: Whether a Graduate Will Migrate to Work

Independent Variables	(1)	(2)	(3)
Study migration	0.3552***	0.3722***	0.3744***
	(0.0309)	(0.0321)	(0.0326)
Age	0.0209	0.0169	0.0124
	(0.0135)	(0.0133)	(0.0133)
Female	-0.061*	-0.0629*	-0.0574*
	(0.0282)	(0.0284)	(0.0285)
Minority	-0.00349	-0.0235	-0.0254
	(0.0508)	(0.0482)	(0.0476)
Unhealthy	0.0708	0.0566	0.0688
	(0.0795)	(0.0759)	(0.0774)
In a relationship	-0.018	-0.0254	-0.0275
	(0.0273)	(0.027)	(0.0269)
Risk appetite	-0.0228	-0.0233	-0.0207
	(0.0151)	(0.0150)	(0.0151)
Only child	-0.0379	-0.0324	-0.0370

	(0.0323)	(0.0326)	(0.0325)
From rural area	0.0458	0.046	0.0396
	(0.0369)	(0.0368)	(0.0370)
SES score	0.0167	0.0239	0.0215
	(0.020)	(0.020)	(0.0208)
Science or engineering major	0.0939*	0.111*	0.1102**
	(0.0453)	(0.046)	(0.0455)
Economics or management major	-0.0323	-0.0268	-0.0308
	(0.0495)	(0.050)	(0.0504)
Student leader in high school	0.0698*	0.0704**	0.0745**
	(0.0273)	(0.0272)	(0.0274)
Humanities track in high school	0.009	0.0092	0.0123
	(0.046)	(0.0466)	(0.0464)
Art & sports track in high school	-0.163*	-0.123	-0.1230
	(0.0626)	(0.0715)	(0.0708)
NCEE score (rescaled to 1-100)	-0.00422	-0.00247	-0.0019
	(0.0024)	(0.0026)	(0.0026)
Average course score in college	0.0024	0.0033	0.0021
	(0.0024)	(0.0025)	(0.0024)
Degree of preference towards one's major	0.0096	0.0056	0.0026
	(0.0186)	(0.0184)	(0.0184)
Passed CET4	0.0974*	0.0956*	0.1001*
	(0.039)	(0.039)	(0.0395)
Passed CET6	0.126**	0.138**	0.1434**
	(0.046)	(0.047)	(0.0473)
Student leader	-0.0216	-0.0236	-0.0193
	(0.034)	(0.0331)	(0.0334)
CCP member	0.0056	0.0162	0.0112
	(0.0304)	(0.0302)	(0.0306)
Has professional certificates	0.0620*	0.044	0.0457
	(0.0261)	(0.026)	(0.0261)
Institution located in Central or West China	0.215***	0.032	-0.0285
	(0.0257)	(0.0386)	(0.0455)
Comprehensive institution	-0.054	-0.0139	0.0179
	(0.042)	(0.047)	(0.0522)
Engineering-concentrated institution	0.256***	0.24***	0.2309***
	(0.0375)	(0.0383)	(0.0376)
985' institution	0.294***	0.280***	0.2830***
	(0.046)	(0.049)	(0.0515)
211' institution	0.036	0.0703	0.0880**
	(0.0297)	(0.0336)	(0.0343)
Independent college	0.031	-0.0089	0.0336

	(0.070)	(0.067)	(0.0710)
Unemployment rate		-0.007	0.0090
		(0.0168)	(0.0189)
GDP		-0.0000671***	-0.0000666***
		(8.31e-06)	(9.07e-06)
Population		-0.0000195***	-0.0000359**
		(5.83e-06)	(12.6e-06)
Provincial size		-0.00164**	-0.0021628***
		(0.00056)	(0.00066)
CPI		0.0554	-0.0126
		(0.0543)	(0.0584)
Sea coastal province dummy			-0.0139
			(0.0395)
ECI			-0.3181**
			(0.1151)
Number of higher education institutions			0.0015
			(0.0011)
N	3543	3543	3543
Wald chi(2)	689.35***	694.93***	702.17***
Pseudo R2	0.2615	0.2828	0.2869

Note: 1. Marginal effects are reported; Robust standard errors in parentheses;
* p<0.05, **p<0.01, *** p<0.001
2. Missing dummies were included in regressions.

Section 5.2 Multinomial Logistic Model

As discussed previously, when considering how students decide between the three options of college stayer, return migrant, repeat migrant, the probit model or binary logit model is no longer suitable. Researcher needs to distinguish the characteristics of the individual and the regional characteristics (Faggian, McCann, & Sheppard, 2006). However, since our model is invariant across alternatives (i.e. we only have case-specific variables to work with), we used a multinomial logit model of the following form:

$$P_{im} = Pr(y_i = m|X_i) = \frac{\exp(\alpha_m + \beta_m X_i)}{\sum_{k=1}^{3} \exp(\alpha_k + \beta_k X_i)}, \qquad \text{m= 1, …, 3,} \qquad (5.1)$$

Where P_{im} is the probability that individual i chooses migration status m= [college stayer, return migrant, repeat migrant] and X_i denotes the regressor matrix, including personal characteristics, and β_m are the parameters to be estimated. Based on utility maximization theory, the multinomial logistic model treats a college graduate's decision (discrete choice situation) as a comparison between the utilities (continuous latent variable) of alternative migration types. The multinomial logic model is a popular framework for estimating the determinants of location choice for immigrants and migrants (Scott, Coomes & Izyumov, 2005). Multinomial logistic regression does not assume normality, linearity, or homoscedasticity; however, it does have assumptions, such as independence of irrelevant alternatives (IIA assumption). This assumption states that the choice of one category within the dependent variable is not related to the choice of another category. The assumption of independence will be tested with the Hausman-McFadden test in this section. The results for multinomial logit model of migration status choice, between (1) college stayer, (2) return migrant, and (3) repeat migrant are presented.

Applying the multinomial logistic regression model to our dataset, there are mainly two parts of results. This breaks the regression up into a series of binary regressions. Stata has the ".mlogit" command for the multinomial logit model. In this study, the college stayer (category 1 of the dependent variable) was used as the base category or comparison group for the estimation.

Migration status is the response variable in the multinomial logistic regression. Underneath migration status are two replicates of the predictor variables, representing the two models that are estimated: return migrant relative to college stayer, and repeat migrant relative to college stayer. The college stayer group is the reference group in this analysis. In statistics, odds are defined as the ratio of the probability of success and the probability of failure. Probability ranges from 0 and 1 while odds range between 0 and positive infinity. The transformation from odds to

log of odds is the log transformation. There are mainly two reasons for conducting the log transformation in practice. First, since probability has a restricted range, it is usually difficult to model a variable's probability. This transformation is an attempt to avoid the restricted range problem. "It maps probability ranging between 0 and 1 to log odds ranging from negative infinity to positive infinity" (UCLA Statistical Consulting Group, 2014). Another reason is that the log of odds is one of the easiest to understand and interpret among all types of transformation.

To obtain the coefficients on the odds ratio scale we just add the option ", rrr" command together with ".mlogit" in Stata to obtain the relative risk ratio (RRR) results. Or, we could exponentiate the coefficient—the log odds from the logistic multinomial regression to get the odds ratio. As shown in Table 5-3, there are two sets of results in the output. In fact, the first set of analyses is as though we performed a simple binary logistic regression where the repeat migrants were omitted. If we did that analysis we would see that the odds ratio from the logistic command would be very similar to the RRR from the ".mlogit" command. Similarly, the second set of analyses is as though we performed a simple binary logistic regression where the return migrants were omitted. And the odds ratio from the logistic command will be consistent with the RRR results from the ".mlogit" command. Therefore, RRR in the first set of multinomial regression results refers to the odds of the return migrant vs. college stayer, and RRR in the second set refers to the odds of the repeat migrant vs. college stayer. The interpretation of the odds ratio is analogous to logistic regression and it is the ratio of two odds, or the change in odds in the multiplicative scale for a unit increase in the corresponding predictor variable holding other variables at a certain value (UCLA Statistical Consulting Group, 2014).

The results indicate the number of observations used in the multinomial logistic regression is 3179 and the log likelihood is -2064.23. The LR chi-square (104) = 1915.11 is the Likelihood

Ratio (LR) Chi-Square. This is the test for both equations (return migrant relative to college stayer and repeat migrant relative to college stayer); at least one of the predictors' regression coefficients is not equal to zero. The findings show a small p-value from the LR test, <0.00001, which would lead us to conclude that at least one of the regression coefficients in the model is not equal to zero. The McFadden's pseudo R-squared is 0.3169.

Table 5-2 Multinomial Regression on Work Migration

College stayer as base outcome	**Return Migrant** Odds Ratio	**Repeat Migrant** Odds Ratio
Study migration	N/A	3.5593***
		(0.4155)
Age	1.0856	1.0873
	(0.0839)	(0.0523)
Female	1.0005	0.6924***
	(0.1764)	(0.0771)
Minority	0.9302	0.7641
	(0.2491)	(0.1474)
Unhealthy	1.3503	1.0542
	(0.5418)	(0.2750)
In a relationship	0.8340	0.8186*
	(0.1318)	(0.0799)
Risk appetite	0.9312	0.9603
	(0.0787)	(0.0502)
Only child	1.1532	0.9721
	(0.2113)	(0.1165)
From rural area	0.790	1.0982
	(0.1649)	(0.1366)
SES score	1.0759	0.9510
	(0.1215)	(0.0687)
Science or engineering major	1.2338	1.7237**
	(0.3473)	(0.3068)

Economics or management major	1.1821	0.9750
	(0.3226)	(0.174)
Student leader in high school	1.6334***	1.1699
	(0.2514)	(0.1127)
Humanities track in high school	0.9289	0.8131
	(0.2463)	(0.1420)
Art & sports track in high school	0.4531	0.5655
	(0.2279)	(0.1868)
NCEE score (rescaled to 1-100)	1.0147	1.0083
	(0.0153)	(0.0102)
Average course score in college	1.0238	1.0090
	(0.0155)	(0.0092)
Preference degree towards one's major	1.0125	1.1561*
	(0.0965)	(0.0708)
Passed CET4	1.9786**	1.5413***
	(0.5204)	(0.2099)
Passed CET6	1.888*	1.5602**
	(0.5485)	(0.2486)
Student leader	0.7110	0.9811
	(0.1412)	(0.1176)
CCP member	0.8786	0.8443
	(0.1528)	(0.0924)
Has professional certificates	1.226	1.0102
	(0.1877)	(0.0961)
Institution located in Central or West China	0.3246**	0.6993
	(0.1289)	(0.1466)
Comprehensive institution	0.7966	1.1575
	(0.2813)	(0.2476)
Engineering-concentrated institution	2.7235***	2.6931***
	(0.8036)	(0.4608)
985' institution	2.7904***	2.7896***
	(0.890)	(0.5671)
211' institution	1.2160	1.2976
	(0.2592)	(0.2794)
Independent college	2.1004	1.9519
	(1.4347)	(0.7405)

Unemployment rate	1.3347*	1.3426***
	(0.1860)	(0.111)
CPI	1.8108	0.8343
	(0.9146)	(0.205)
GDP	0.9997***	0.9998***
	(0.00006)	(0.00004)
Population	1.000004	0.9999
	(0.0001)	(5.62e-05)
Provincial size	0.9822***	0.9896***
	(0.0054)	(0.0028)
Sea coastal province dummy	0.8097	0.6110**
	(0.2668)	(0.1069)
ECI	0.0994***	0.0678***
	(0.0713)	(0.0312)
Number of higher institution	0.9889	1.0013
	(0.0085)	(0.0053)
Constant	2.23e-36	1.26e+07
	(1.18e-33)	(3.20e+08)
N	3179	3179
Chi-squared	1915.11	1915.11
Pseudo R-squared	0.3169	0.3169

Note: 1. Regression coefficients are reported; Standard errors in parentheses;
2. * p<0.05, **p<0.01, *** p<0.001
3. Missing dummies were included in regressions.

1) Return migrant relative to college stayer

Student leader in high school—The coefficient reported in Table 5-3 for "student leader in high school" is the odds ratio comparing college graduates who are student leaders in high school to those who are not student leader in high school given that the other variables in the model are held constant. This can be interpreted as follows: holding other covariates at a fixed value, the odds of getting into the return migrant group (instead of being in the staying group) for

those who are student leaders in high school over the odds of getting into the return migrant group (instead of being in the staying group) for those who are not student leader in higher school is 1.6334. In terms of percent change, we can say that the odds of being in the return migrant group for students who are leaders in high school are 63.34% higher than the odds for students who are not leaders in high school.

Being a student leader in high school is an indicator of accomplishment for a person. College graduates who were high school leaders might feel more fulfilled and content in their hometown, which may give impetus to their moving back to hometown.

CET4— Given that the other variables in the model are held constant, the odds of getting into the return migrant group for those who passed the CET4 over the odds of getting into the return migrant group for those who did not pass the CET4 is 1.9786. Put another way, the odds of being in the return migrant group for students who passed the CET4 are 97.86% higher than the odds for students who did not pass the CET4.

CET6—Similarly, the odds of being in the return migrant group for students who passed the CET6 are 88.8% higher than the odds for students who did not pass the CET6 given that the other variables in the model are held constant.

Institution is located in Central or West China— The odds of being in the return migrant group for those students who are from institutions located in Central or West China over the odds for students from institutions located in other regions is 0.3246. In other words, the students who are from institutions located in Central or West China are less likely to be in the return migrant group relative to the college stayer group. McHugh (2009) suggested that when the odds of the first group experiencing the event are less than the odds of the second group, one must reverse

the two columns so that the second group becomes the first and the first group becomes the second[11]. In this way it will be possible to interpret the difference because that reversal will calculate how many more times the second group experienced the event than the first. To obtain the percentage interpretation, the model was run again using the return migrant as reference group. The coefficient for variable "institution located in Central or West China" is 3.0805. This indicates that the odds of being in the college stayer group for students who are from institutions located in central or western regions over the odds for students who are not is 3.0805. The odds of getting in the college stayer group for students who are from institutions located in central or western regions are 208.05% higher than the odds for students from institutions located in other regions, holding other variables in the model constant. College graduates from institutions located in Central or West China would become college stayers 3.08 times more often than graduates from institutions located in other regions.

Institution is engineering concentrated—The odds of getting in the return migrant group for college graduates from engineering-concentrated institutions over the odds for students from institutions with other concentration is 2.72, ensuring other variables in the model are held constant. In terms of percentage wise, the odds of being in the return migrant group for students who are from engineering-concentrated institutions are 172.35% higher than the odds for students who are not from engineering-concentrated institutions.

The 985 institutions--Assuming other variables in the model are held constant, students from 985 institutions are expected to become return migrants 2.79 times more often than those who are not from 985 institutions. In percent wise, the odds of getting into the return migrant group

[11] According to McHugh (2009), "Odds ratio of less than 1 means that the first group was less likely to experience the event. However, an OR value below 1 is not directly interpretable. The degree to which the first group is less likely to experience the event is not the OR result….It is not valid to try to determine how much less the first group's odds of the event was than the second group's. "

for students who are from 985 institutions are 179.04% higher than the odds for those who are not from 985 institutions.

Unemployment rate—For continuous variable, each coefficient is the ratio of two odds, or the change in odds in the multiplicative scale for a unit increase in the corresponding continuous predictor variable holding other variables at certain value. Therefore, with the other variables in the model held constant, we will see a 33.47% increase in the odds of getting into the return migrant group for a one-unit increase in the provincial unemployment rate.

Gross domestic product, provincial size, and ECI all have a statistically significant impact on work migration. The coefficients for these three variables are less than one, therefore, the model using return migrant as the reference group was run so that it was possible to interpret the difference in odds. If there is a one dollar increase in provincial GDP per capita, the odds of being in the college stayer group will increase 0.03% given that the other variables in the model are held constant. So, given a one dollar increase in provincial GDP per capita, the odds of being in the college stayer group would be 1.0003 times more likely when the other variables in the model are held constant. More specifically, we can say that if a student's institution is located in higher GDP province, the student would be more likely to fall into the college stayer group than the return migrant group. Similarly, if there is a 10,000 square kilometer increase in provincial area size, the odds for college stayer relative to repeat migrant would be expected to increase by 1.814% given that the other variables in the model are held constant. More specifically, we can say that if a student's institution is located in one of the larger provinces, the student would be expected to fall into the college stayer group rather than the repeat migrant group. Given a one-unit increase in ECI score, we will see 906% increase in the odds of being in the college stayer group holding other variables at the fixed values. We can say that if a student's institution is

located in higher ECI province, the student would be more likely to fall into college stayer group as opposed to the return migrant group.

2) Repeat migrant relative to college stayer

Study migration— The odds of getting into the repeat migrant group for students who migrated for college are 255.93% higher than the odds for those who did not migrate for college given that other variables in the model are held constant. In other words, if a student migrated to attend college, they would be more likely to fall into the repeat migrant group rather than the college stayer group.

Science or engineering major—Students with a science or engineering major is a statistically significant factor in predicting work migration. Holding other variables at a fixed value, the odds of getting into repeat migrant group (instead of being in the staying group) for students with a science or engineering major over the odds of getting into repeat migrant group (instead of being in the staying group) for students with other major is 1.72. In terms of percent change, we can say that the odds for students with science or engineering major are 72% higher than the odds for students with other major.

Degree of preference towards one's major—If there is a one-unit increase in the degree of preference towards one's major, the odds for a student being a repeat migrant relative to college stayer would be expected to increase by 15.61% assuming all other variables in the model remain constant.

CET4— Given that the other variables in the model are held constant, the odds of getting into the repeat migrant group for those who passed the CET4 over the odds of getting into the college stayer group for those who did not pass the CET4 is 1.5413. The odds of being in the

repeat migrant group for students who passed the CET4 are 54.13% higher than the odds for students who did not pass the CET4.

CET6—Similarly, the odds of being in the repeat migrant group for students who passed the CET6 are 56.02% higher than the odds for students who did not pass the CET6 given the other variables in the model being held constant.

Institution is engineering concentrated—The odds of getting into the repeat migrant group for college graduates from engineering-concentrated institutions over the odds of getting into the repeat migrant group for students from other concentration institutions is 2.69, given other variables in the model being held constant. In terms of percentage wise, the odds of being in the repeat migrant group for students who are from engineering-concentrated institutions are 169% higher than the odds for students who are not from engineering-concentrated institutions.

The 985 institutions—The odds of getting into the repeat migrant group for students from 985 institutions over the odds of getting into the repeat migrant group for students who are not from 985 institutions is 2.79, given other variables in the model being constant. In percent wise, the odds of getting into the repeat migrant group for students who are from 985 institutions are 179% higher than those who are not from 985 institutions.

Unemployment rate—With the other variables in the model held constant, we will see a 34.26% increase in the odds of getting into the repeat migrant group for every one-unit increase in the provincial unemployment rate.

In order to interpret the odds ratio less than 1, a regression was run using repeat migrant as the comparison group.

The fitted model shows that holding other variables at fixed values, the odds of getting into college stayer group for females over the odds of getting into college stayer group is 1.4442. In terms of percent change, we can say that the odds of getting into the college stayer group for female students are 44.42% higher than male students. The odds of getting into the college stayer group for students who reported being "in a relationship" are 22.16% higher than those who are not.

If there is a one-dollar increase in provincial GDP per capita, the odds of being in the college stayer group will increase by 0.02% given that the other variables in the model are held constant. More specifically, we can say that if a student's institution is located in a province with a higher GDP, the student would be more likely to fall into the college stayer group rather than the repeat migrant group.

If there is a 10,000 square kilometer increase in provincial size, the odds for being a college stayer relative to a repeat migrant would be expected to increase by 1.05% given that the other variables in the model are held constant. More specifically, we can say that if a student's institution is located in a province with a larger land area size, the student would be more likely to fall into the college stayer group as opposed to the repeat migrant group.

Holding other variables in the model at fixed values, we will see 1376% increase in the odds of getting into the college stayer group for a one-unit increase in ECI score. We can say that if a student's institution is located in a province with higher ECI score, the student would be predicted to go into the college stayer group rather than the repeat migrant group.

The Hausman test was employed to test the IIA assumption in this study. In Stata, ".mlogtest" command was used to conduct the Hausman test for the multinomial regression's IIA

assumption. All the values of chi-square for the three categories college stayer, return migrant, and repeat migrant are not negative, which means the estimated model does meet the asymptotic assumption of the test. Also, the results show that the tests are "for H_0", in other words, the relative probability of choosing to be return migrant over being a college stayer is independent of the option of being a repeat migrant. The independence of the irrelevant alternatives assumption is satisfied.

Section 5.3 Summary of Empirical Findings and Discussion

This chapter presents the analysis of the intention-to-work sample to answer the first research question—What are the determinants of college graduates' work migration decision in China? The probit model was first employed when the two options "migrate" and "do not migrate" were taken into consideration. The multinomial model was then estimated to deal with the three choices of "college stayer", "return migrant" and "repeat migrant".

The results show that the following variables have a significant positive impact on work migration: study migration, science or engineering major, student leader in high school, passed CET4, passed CET6, engineering-concentrated institution, from 985 institutions, and from 211 institutions. College graduates who possess the above individual characteristics or are from the above institutions are more likely to migrate to work.

The following variables have a significant negative impact on work migration: female, provincial GDP per capita, provincial population, provincial area size, and provincial ECI score.

It was found that previous migration behavior has a statistically significant positive impact on work migration. Having migrated to college before is associated with an increase of about 37% in the probability of work migration. This finding is consistent with previous research that found

prior migration is highly correlated with subsequent migration (Adelman, 2004; Kodrzycki, 2001; Perry, 2001; Tornatzkey et. al, 1998, 2001).

Unlike previous research results, age was not a significant variable influencing migration in this study. The reason for this is that all the survey objects were college students of a similar age. Therefore, there was not enough variation in age for the sample to reflect the importance of age on work migration.

Consistent with some previous studies (Faggian, McCann & Sheppard, 2007), this study found out that male students are statistically significantly more mobile than female students. The explanation may be that men tend to be more committed to their careers than women, and therefore men are more likely to make the necessary moves required in order to achieve career development or a good career start in this case (Faggian, McCann & Sheppard, 2007). Because of data limitations, the scenario of migration after women rear children cannot be examined. College graduates who were student leaders in high school, had passed CET4, and passed CET6 may indicate students have higher ability and they tend to be more mobile than their peers.

Institutional characteristics also play an important role on the work migration decision. If one is majoring in science or engineering major, or attending an engineering-concentrated institution, one is more likely to migrate to work. The reason for this finding might be that in China large-scale industries are distributed all over the country and engineering students are highly needed by these factories and companies. Therefore they are more likely to move for such jobs.

The results also showed that provincial GDP per capita, population, size, and ECI all have a significant impact on work migration. The larger the provincial area size or population is (where

the higher institution is located), the less likely the student is to migrate to work. The higher the provincial GDP per capita or ECI score is, the less likely the student is to migrate to work. In another word, the regression analyses indicated that students tended to move away from provinces with poorer economic attributes and environment while preferring to remain in larger provinces. The findings again confirm that economic opportunities and amenities influence the location decisions of college graduates. Differential characteristics of regions (e.g. regional size, general labor market, prevailing conditions in land and housing markets, interregional differences in regional salaries and regional employment opportunities, cultural and social environment, etc.) have important impact on moving (Kyung, 1992; Abbott & Schmid, 1975, Gossman et al., 1968).

.

CHAPTER SIX:
IMPACT OF MIGRATION ON GRADUATES' STARTING SALARIES

This chapter presents the empirical findings for the second research question about impact of work migration behavior on graduates' early post-college labor market performance—starting salary. Different models based on the Mincer earnings function will be employed to examine this research question. There might be several sources of bias associated with the basic model. Though the College Students' Labor Market survey data includes extensive information about the student, it is still likely that some important determinants (e.g., unobserved ability) of starting salary are not included in the model. In other words, college graduates' migration behavior is possibly correlated with some unobservable student characteristics such as individual ability. Omission of such variables, particularly when they are also correlated with included explanatory variables will cause bias in identifying the link between initial salary and the included variables. Apart from the omitted variables, another source of bias may come from the endogeneity of migration. Roy (1951) indicated that migration is endogenous and college graduates self-select into migration. In other words, the causality between a graduate's initial salary and his/her migration behavior could be reversed.

To make a more accurate investigation of the causal effects of migration on graduates' starting monthly salary, this study addresses the endogeneity problem with two quasi-experimental strategies. The basic idea is to construct a comparable control group which is similar to the treatment group in every observed aspect except for the treatment status (work migration). The two identification strategies employed are Propensity Score Matching (PSM) and Instrumental Variable (IV) design. Propensity score matching provides a means for adjusting for selection bias in observational studies of causal effects by matching different groups based on

their propensity score probabilities. Instrumental variable is another identification strategy to be used to make possible estimates of causal effects without random assignment.

In Section 6.1, the results of the probit model on initial employment status are reported and the Heckman test for sample selection bias is employed. The OLS estimates for impact of work migration on graduates' starting monthly salary are reported in Section 6.2. The findings based on the PSM model and the IV models are presented in Section 6.3 and Section 6.4 respectively. Section 6.5 gives a summary of the findings for the research question about the impact of work migration on starting salary. For simplicity, the coefficients of missing value dummy variables are not presented for any of the models.

Section 6.1 Probit Model on Initial Employment Status

This study only uses a subgroup of college graduates—a group of students who have an intention to work after graduation (intention-to-work sample) and have already secured job offers with observable salary values. This may be a non-random sample of all college graduates who intend to work, if there are variables which affect college graduates getting job offer in the labor force. If so, then OLS estimates for the sample of graduates who had offers will be biased and inconsistent.

The Heckman correction method was originally designed to address the problem of estimating the average salary of women using data collected from a population of women in which housewives were excluded by self-selection. Heckman's model lays the groundwork for understanding the treatment effect model and is among the most important contributions to program evaluation (Guo & Fraser, 2009).

The Heckman correction method involves both the estimation of a selection equation (probit model) and a second outcome equation (insertion of a correction factor—the inverse Mills ratio, calculated from the probit model—into the second OLS model of interest). It is used to assess whether selection is a problem. In this study, the first step is the estimation of the probability of having an offer as a function of the original control variables and an additional identifying variable. The second step is the estimation of the Mincer equation for starting salary based on OLS. Each step has a residual for each observation, or a set of unknowns for each observation. To test for bias, we examined the relationship between the residuals for the two steps. If the unobservables in the selection model are correlated with the unobservables in the step 2 model, we have biased estimates without correction (or in an OLS model). This is basically saying that unobservables in the selection of employment are also affecting the step 2 model. If the unobservables in step 1 are unrelated to the unobservables in step 2, then we are saying that selection into the step 2 sample is a random process, unaffected by additional unobservables. The probability of us having selection bias is small.

In this study, two variables, "whether the institution is in a small city" and "number of job applications submitted", were added to the equation in the first step. Each variable was used as an identifying variable to conduct the Heckman test. These two variables are assumed to affect the probability of getting an offer, but do not influence the starting monthly salary.

The first identifying variable is whether the institution is located in a small city. Small cities are assumed to have fewer job opportunities and the institutions might be less attractive for big companies' campus recruiting. Students may have to travel to nearby big cities to attend job fairs or have job interviews, which may increase the difficulties for students in small cities to get job offers before graduation. As with the results shown in Table 6-1, the coefficient on the variable

"whether institution is located in a small city" in the probit model is significant and negative. However, the institution's location would have no influence on the monthly starting salary as long as the industry type, position, and location of workplace are controlled for. When this variable is included in the salary equation, the coefficient is not statistically significant. This verifies that whether the institution is in a small city or not has no direct impact on starting monthly salary.

Another identifying variable is the number of job applications submitted by the student. It might seem obvious that students who submit more job applications could potentially get more job offers, from which they would be able to pick one with a higher salary. However, the salary is mainly dictated by other factors such as the nature of the job, position, industry, and location of workplace. When this information is considered, the only path that the number of job applications submitted by individual students would have impact on the salary is through its impact on whether the student might get a job offer. Therefore the exclusive condition is satisfied. When added to the salary equation, the coefficient on this number of job applications submitted variable was not statistically significant. This verified that the number of submitted job applications had no direct impact on college graduates' starting salaries. As shown in the Table 6-1 (which shows the probit results from the first step of the Heckman correction model), the assumptions have been tested.

Consider the results in Table 6-1. Being the only child in one's family is statistically significantly associated with a lower probability of getting an employment offer. If the student is a student leader at high school, then the student is significantly more likely to be offered a job. As for the student's academic achievement, students who have higher average score in college are less likely to get an employment offer, though the impact is very small. A one point increase

in average course score is associated with a decrease of 0.7% in the probability of being offered a job. As for English proficiency, neither coefficients for passing CET4 and CET6 are statistically significant.

Among college experience variables, a student's major significantly influences their employment status. Compared with students with humanities majors, graduates with a science or engineering major have a higher probability of being employed. Also, preference of degree towards one's major is a significant predictor of obtaining an offer. Whether students have ever worked in college is a significant predictor of getting an employment offer. In terms of institution characteristics, both academic ranking and institution concentration have a significant impact on graduates' labor market outcomes. Students from 985 institutions have a higher probability of being employed than those in non-key institutions. Students from independent colleges are significantly less likely to get an offer. Compared with students in institutions with other concentrations, students from engineering-concentrated institutions are significantly more likely to be offered a job.

Other variables including age, minority, rural, SES score, high school academic track, whether student has a minor, student leader in college, whether student is a CCP member, whether student has professional certificates, type of financial aid, and institution in central and west areas are not significant predictors of future employment offers.

Stata has the ".pweight" code to specify probability weights. When computing estimates such as standard deviation and regressions parameters, the pweight command causes Stata to use the sampling weight as the number of subjects in the population that each observation represents. A robust variance estimation technique will automatically be used to adjust for the design

characteristics so that variances, standard errors and confidence intervals are correct. In the regression analysis in this thesis, pweight code will be used.

Table 6-1 Probit Model on Initial Employment Status
(with intention-to–work sample
Dependent variable: Whether offered a job before graduation)

	With "institution is located in a small city" as exclusive variable	With "number of job application" as exclusive variable
Age	0.001	-0.003
	(0.012)	(0.012)
Female	-0.023	-0.045
	(0.026)	(0.027)
Only child	-0.071*	-0.057*
	(0.028)	(0.028)
Minority	-0.035	-0.048
	(0.044)	(0.045)
From rural area	0.007	0.002
	(0.031)	(0.032)
SES score	-0.017	-0.015
	(0.017)	(0.017)
Science or engineering major	0.103**	0.095**
	(0.038)	(0.040)
Economics or management major	0.021	0.017
	(0.036)	(0.037)
Student leader in high school	0.053*	0.049*
	(0.023)	(0.024)
Humanities track in high school	0.012	0.013
	(0.037)	(0.038)
Art & sports track in high school	0.018	0.012
	(0.051)	(0.054)
NCEE score (rescaled to 1~100)	0.003	0.004
	(0.002)	(0.002)
Average course score in college	-0.007**	-0.006**
	(0.002)	(0.002)
Has a minor	0.030	0.034
	(0.043)	(0.044)

Degree of preference towards one's major	0.038**	0.038**
	(0.015)	(0.015)
Passed CET4	0.046	0.045
	(0.031)	(0.031)
Passed CET6	0.050	0.047
	(0.035)	(0.036)
Student leader	-0.010	-0.006
	(0.029)	(0.03)
CCP member	0.021	0.046
	(0.027)	(0.027)
Has professional certificates	0.032	0.030
	(0.023)	(0.023)
Ever worked in college	0.148***	0.115***
	(0.030)	(0.032)
Had merit-based aid	0.039	0.015
	(0.028)	(0.028)
Had needed-based aid	0.046	0.046
	(0.029)	(0.031)
Had loans	0.044	0.048
	(0.028)	(0.028)
985 institution	0.064	0.102**
	(0.034)	(0.031)
211 institution	-0.028	0.005
	(0.023)	(0.023)
Independent college	-0.182***	-0.114**
	(0.049)	(0.050)
Comprehensive institution	0.032	0.024
	(0.031)	(0.032)
Engineering-concentrated institution	0.163***	0.165***
	(0.029)	(0.030)
Institution located in Central or West China	0.015	-0.010
	(0.024)	(0.025)
Institution located in small city	-0.054*	
	(0.024)	
Number of job applications		0.002**
		(0.001)
N	4917	4917

Pseudo R-square	0.180	0.191

Note: 1. Marginal effects instead of standard coefficients are reported
2. + p<0.10, * p<0.05, ** p<0.01, *** p<0.001
3. Robust S.E in parentheses
4. Missing dummies were included in regressions.

Section 6.2 Impact of Work Migration on Starting Salary, OLS Estimation

As discussed in the previous section, the basic model without selection problem and endogeneity can be written as,

$$InW = \beta_0 + \beta_1 X_i + \beta_2 Z_{ij} + \delta_1 M_{ij} + u_{ij}$$

The dependent variable InW is the log form of the starting monthly salary in RMB. M_{ij} is a measure of work migration, X_i is a set of individual covariates and Z_{ij} is institution characteristics. The model is first estimated with OLS regression. The sample used to estimate the work migration's impact on starting salary were the students who were in the intention-to-work sample, had an offer of employment at the time of the survey and reported the salary offer in the CSLM 2011 survey.

Table 6-2 represents the regression results from several equations. Column 1 presents the basic OLS salary regression results (weighted) of the impact of work migration on starting monthly salary to provide a baseline of comparison[12]; Column 2 presents the OLS results without sampling weights; Column 3 presents the second step of the Heckman test with the number of submitted job applications as the exclusive variable; Column 4 presents the Heckman test with whether or not the institution is located in a small city as the exclusive variable. The number of observations included in the regression was 3146 and the adjusted R-square for the weighted

[12] Results for separate OLS regressions by gender are presented in Appendix 2.

OLS model was 0.328, which means that about 32.8% of the variation in starting monthly salary could be explained by the regression. The adjusted R-square was 0.269 for the OLS regression without sampling weights, which means that about 26.9% of the variation in starting monthly salary could be explained by the OLS regression without sampling weights. OLS estimates without sampling weights are presented here for the purpose of comparison with the PSM estimates in the next section. Because there is no available package in Stata 12 to incorporate sampling weights in the propensity score matching process, therefore, results from an OLS regression without sampling weights are presented as the baseline of comparison.

Table 6-2 OLS Estimates of Impact of Work Migration on Starting Salary (Dependent variable: Starting monthly salary in log form)

	Starting monthly salary			
	OLS (weighted)	**OLS (unweighted)**	**Heckman1 (with smallcity)**	**Heckman2 (with résumés)**
	(1)	**(2)**	**(3)**	**(4)**
Work migration	0.099***	0.0859***	0.0966***	0.0993***
	(0.027)	(0.0181)	(0.0269)	(0.0272)
Age	0.006	0.00549	0.0069	0.0065
	(0.009)	(0.0067)	(0.0092)	(0.0093)
Female	-0.087***	-0.0821***	-0.0932***	-0.0888***
	(0.021)	(0.0155)	(0.0222)	(0.0219)
Minority	0.012	0.0265	0.0028	0.0096
	(0.045)	(0.0263)	(0.0455)	(0.0448)
Only child	0.003	-0.0102	-0.0145	0.0004
	(0.024)	(0.0163)	(0.0308)	(0.0246)
From rural area	-0.038	-0.025	-0.0371	-0.0386
	(0.026)	-0.0172	(0.0261)	(0.026)
SES score	0.021	0.0330***	0.0161	0.0205
	(0.014)	-0.0099	(0.0152)	(0.0141)
Science or engineering major	-0.026	-0.0233	0.0018	-0.0214
	(0.035)	-0.0242	(0.0446)	(0.0361)
Economic or management	-0.126***	-0.0835***	-0.118**	-0.125***

major	(0.036)	-0.0237	(0.0368)	(0.0363)
Student leader in high school	0.041*	0.0272*	0.0541*	0.0436*
	(0.019)	-0.013	(0.0246)	(0.019)
Humanities track	-0.061	-0.0555*	-0.0573	-0.0603
	(0.034)	-0.0225	(0.0346)	(0.0342)
Arts or sports track	-0.003	0.0162	0.0039	-0.0025
	(0.061)	-0.0418	(0.0612)	(0.0607)
NCEE (rescaled to 1-100)	0.007***	0.0058***	0.0078***	0.0072***
	(0.002)	-0.0013	(0.002)	(0.0018)
Average course score	0.000239	0.0008	-0.0017	-0.00008
	(0.002)	-0.0013	(0.0027)	(0.0018)
Has a minor	-0.004	-0.0060	0.0019	-0.0028
	(0.033)	-0.0252	(0.0338)	(0.0334)
Preference degree towards one's major	0.02	0.0252**	0.0298*	0.0215
	(0.011)	-0.0083	(0.0152)	(0.0113)
Passed CET6	0.143***	0.128***	0.158***	0.145***
	(0.031)	-0.0215	(0.0345)	(0.0309)
Passed CET4	0.059*	0.0724***	0.0719*	0.0607*
	(0.026)	-0.0185	(0.0297)	(0.0261)
Student leader	0.033	0.0336*	0.0308	0.0326
	(0.024)	-0.0162	(0.0226)	(0.0226)
CCP member	0.033	0.0453**	0.0382	0.0344
	(0.022)	-0.0148	(0.0224)	(0.0216)
Has professional certificates	-0.012	0.0115	-0.0042	-0.0112
	(0.018)	-0.0127	(0.0196)	(0.0183)
Had merit-based aid	0.039	0.0191	0.0498*	0.0406
	(0.022)	-0.0152	(0.0237)	(0.0217)
Had needs-based aid	-0.032	-0.0225	-0.0205	-0.0298
	(0.021)	-0.0159	(0.0231)	(0.0216)
Had loan	-0.017	-0.0195	-0.0067	-0.0147
	(0.021)	-0.0147	(0.0233)	(0.0208)
Comprehensive institution	0.095*	0.0700**	0.110*	0.0973*
	(0.040)	-0.0243	(0.0435)	(0.0397)
Engineering-concentrated institution	0.042	0.0492*	0.0905	0.0498
	(0.030)	-0.0231	(0.0609)	(0.0318)
985 institution	0.131***	0.168***	0.153***	0.136***
	(0.039)	-0.0259	(0.0422)	(0.0388)
211 institution	0.126***	0.115***	0.123***	0.126***
	(0.020)	-0.0167	(0.0208)	(0.0204)
Institution located in central or west area	0.021	-0.0072	0.0244	0.0209
	(0.022)	-0.018	(0.0227)	(0.0223)
Independent college	-0.069	-0.0903	-0.105	-0.0743
	(0.046)	-0.0523	(0.0628)	(0.0477)

Ever worked in college	-0.028	-0.0473*	0.0102	-0.0227
	(0.026)	-0.0185	(0.05)	(0.0266)
Constant	6.782***	6.861***	6.649***	6.753***
	(0.285)	(0.215)	(0.320)	(0.289)
Lambda			0.197	0.0406
			(0.223)	(0.0592)
N	3146	3146	3146	3146
R-square	0.3278	0.2924	0.3282	0.3281
Adj. R-square	0.306	0.2692	0.306	0.306

Note: 1. + p<0.10, * p<0.05, ** p<0.01, *** p<0.001
2. Robust S.E in parentheses
3. Missing dummies, province dummies, employer dummies and industry dummies are included in the regression, but their coefficients are not reported

According to the weighted OLS results shown above, work migration has a statistically significant positive effect on a student's starting monthly salary. Because the dependent variable is a log of the starting monthly salary, the estimated coefficients shown in the table can be interpreted as marginal effects (in percentage). The coefficient β for migration is the marginal effects of work migration on starting monthly salary for the work migrants group over college stayer group, when other predictors are held at a fixed value. In this case, the coefficient for migration is 0.099, which means that holding all the other covariates constant, new graduates who migrated enjoy a 9.9% starting monthly salary premium over those graduates who did not migrate in China, as shown in the model.

The results also show that students' individual characteristics, family background, academic achievement, college experience and institution attributes all influence their early post-college labor market outcomes. In details, the expected starting monthly salary will be 8.7% lower for female graduates than their male counterparts based on the results of the regression Equation (1). Students with an economics and management major will get a salary 12.6% lower than students

with a humanities major. If the student was a student leader in high school, his/her starting monthly salary will be 4.1% higher than those who were not. For the NCEE standard score, we can say that for a ten-unit increase in NCEE, we would expect to see about a 7% increase in starting monthly salary.

Graduates who passed the CET6 test are likely to have a salary 14.3% higher than those who did not pass the test. Students who passed the CET4 test could potentially have a salary 5.9% higher than those who did not pass the test. If a student graduated from a comprehensive institution, he/she will have a salary about 9.5% higher than those from institutions with other special concentrations other than engineering. Compared with graduates from non-key higher institutions, graduates from 985 institutions will get a 13.1% higher salary. Compared with graduates from non-key institutions, graduates from 211 institutions will get a 12.6% higher salary. The intercept becomes less interesting when some of the predictor variables are not centered and are continuous.

Column 2 of Table 6-2 represents the OLS results without sampling weights. In this regression equation, work migration also has a significant positive effect on a student's starting monthly salary. When other predictors are held at a fixed value, the average starting monthly salary will be about 8.59% higher for the work migration group over the college stayer group. Similarly, the expected starting monthly salary will be 8.21% lower for female graduates than male graduates holding other variables constant. In this model, the SES score is a significant predictor of starting monthly salary; for a one-unit increase in SES score, we would anticipate a 3.3% increase in starting monthly salary. Similar to the results from Equation (1), students with an economics and management major will get a salary 8.35% lower than students with other majors. As for the high school academic track, students on a humanities track will have a 5.56%

lower salary than students on other academic tracks. If the student was a student leader in high school, his/her starting monthly salary will be 2.72% higher than those who were not. For the NCEE standardized score, we can say that for a ten-unit increase in NCEE score, we would expect to see a 5.77% increase in starting monthly salary.

Students who like their major will have a starting monthly salary 2.52% higher than other students. Graduates who passed the CET6 test will have a salary 12.8% higher than those who did not. Students who passed the CET4 test will have a salary 7.24% higher than those who did not. If a student graduated from a comprehensive institution, he/she will have a salary about 7% higher than those from institutions with special concentration other than engineering. Graduates from institutions with an engineering concentration will get a salary 4.92% higher than students from institutions with other special concentration. Compared with students who were not student leaders in college, those who were are likely to have a salary 3.36% higher. The salary for students who are CCP members is 4.53% higher than those who are not CCP members. Compared with graduates from non-key higher institutions, graduates from 985 institutions will get a 16.8% higher salary. Compared with graduates from non-key institutions, graduates from 211 institutions will obtain an 11.5% starting monthly salary premium. Students who have worked in college will have a salary 4.62% lower than those who have not worked.

The column 3 and column 4 in Table 6-2 present the salary models with the two exclusive variables respectively; they are the second stage regression models for the Heckman correction technique. The regression results from the Heckman tests are quite consistent with the OLS regression results. From the results shown, the Lamda for the identifying variables is not statistically significant, which means that unobservables in the selection of employment are unrelated to the unobservables in the step 2 Heckman salary equation. In other words, this

verifies that the number of submitted job applications and whether or not the institution is located in a small city has no direct impact on college graduates' starting salary, suggesting both variables satisfy the exclusive condition. Therefore, we can say that selection into the wage equation sample is a random process, unaffected by additional unobservables. The probability that we will have selection bias is very small.

Section 6.3 Impact of Work Migration on Starting Salary, Propensity Score Matching Method

One of the techniques to resolve the endogeneity problem is to employ the PSM strategy. If students with higher inner ability are more likely to migrate for work, the OLS estimates would be upward biased; if students with lower inner ability are more likely to migrate, then the OLS estimates would be downward biased. In this section, the PSM method will be employed and the results will be compared with those from the OLS.

Propensity score matching is used to match treated and untreated observations on the estimated probability of being treated (propensity score). Many characteristics will be used in the matching process to construct a comparison group (no migration) that is similar to a treatment group (migration) along those characteristics. Observations in the matched group have the same or a very close probability of being treated but are different in the actual treatment status. This section first presents the process of the construction of the propensity score, then it describes the matching process and checks the validation of the two assumptions after matching, and finally it shows the PSM estimates and compares the results with the OLS estimates.

The first step of the PSM strategy is to estimate the probability of work migration with all the available covariates. In order for the propensity scores to correctly estimate the probability of participation, the characteristics included in the propensity score estimation should be as

exhaustive as possible. However, those characteristics which may have been affected by the treatment should not be included. In this study, it is important to include as many confounding factors that predict both migration and the starting salary as possible. These confounding covariates are the ones that we care about to balance across groups. A probit regression will be run where the dependent variable is 'work migration' and the predictors are all the confounding covariates. The probit equation will then be used to compute predicted probabilities for each person that they received the treatment—these are the propensity scores.

Specifically, the model includes individual and family demographic characteristics such as age, gender, and race, whether the student is an only child, the household's SES score, whether the student is from a rural area, whether the student is from Central or Western China, and family income. The variable, study migration is also added in the model due to its significant impact in the previous probit model to predict work migration. For measures of student ability, NCEE scores, whether the student was a leader in high school, average course scores in college, and whether the student passed the CET4 and CET6 are included in the model to estimate the propensity score. The two variables, whether student is on humanities track in high school and whether student is on arts or sports track in high school are included. The model also incorporates the covariate "degree of preference towards one's major". As for college experience, academic major, whether the student is a student leader, whether the student is a CCP member, whether the student received merit-based aid, needs-based aid or had a loan, and whether the student have worked at college are included in the model as well. With regards to institutional characteristics, the model includes the academic ranking level, concentration, region, campus location, and the percentage of work migration students. The model also incorporates the GDP, population and size information of the province where the institution is located. Some

interactions between variables and quadratic forms are added to the model to achieve a better balance. The model is estimated with probit regression via Stata. The sampling weight is not applied in estimations because there is no available package in Stata 12 to incorporate sampling weights in the PSM process. Therefore, the PSM estimates will only be comparable to the OLS and IV regression estimates without sampling weights. In this section, results from an OLS regression without sampling weight are presented as the baseline of comparison. The estimates results from the PSM strategy will be used to examine the direction of the bias in the OLS estimates.

The matching is performed with students in the "have-salary" sample. The *psmatch2* procedure shows that all observations are on the common support in the sample and Figure 6-1 below represents the distribution of the p-score of treated and untreated groups of the have-salary sample. According to the common support condition, the estimated propensity score should be bounded between 0 and 1 and have sufficient overlaps between the treated and untreated groups so that observations can be matched up (Hirono & Imbens, 2001). Figure 6-1 presents the distribution of the propensity score for students who migrated (the treated group) and students who did not migrate to work (the untreated group) to test this condition. The figure shows the treated cases in red on top and the control cases in blue on bottom. The graph clearly shows that lots of control cases have propensity scores close to zero and there seem to be fewer control cases with propensity scores greater than 0.8. Meanwhile, there are treatment cases everywhere, and some of them appear to be concentrated above propensity scores of 0.4. A histogram of the control cases with propensity scores greater than 0.4 (Figure 6-2) was created to examine whether there was a common support problem. From the plot, it is clear that there are control cases that span the full range of propensity scores, though at some points there are fewer cases.

In addition, the study will apply matching with replacement, which means that a control case will be used over and over again as it is the best match for treated cases. In summary, the figure suggests that all observations are on the common support, and there are sufficient overlaps in the propensity score between the treated and untreated groups.

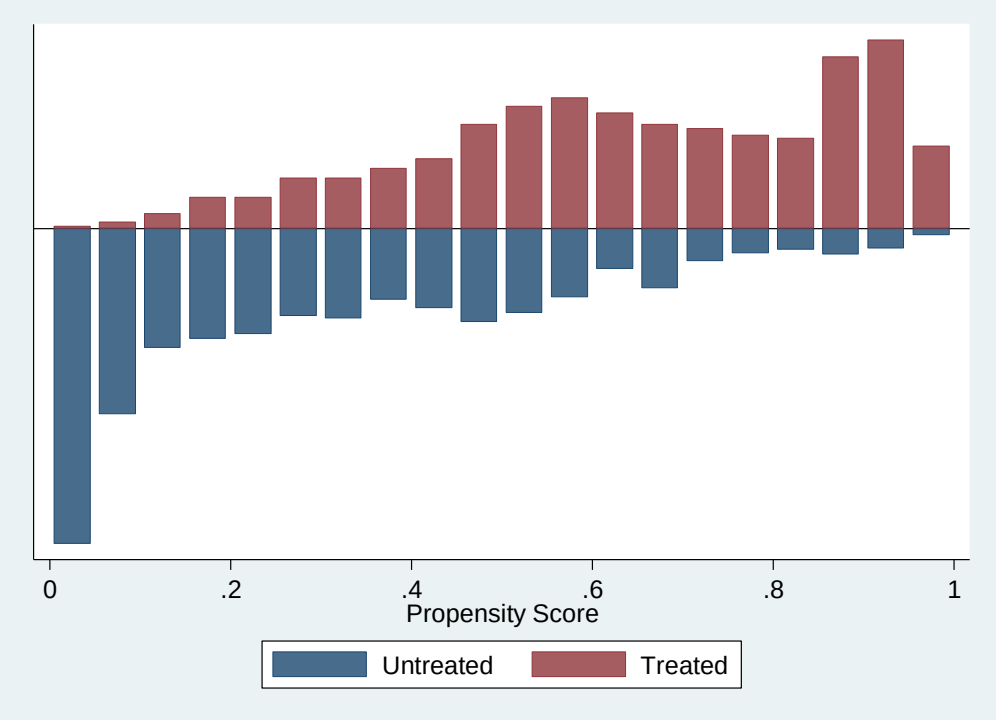


Figure 6-1 Distribution of the p-score of Treated and Untreated Groups of the "Have Salary" Sample

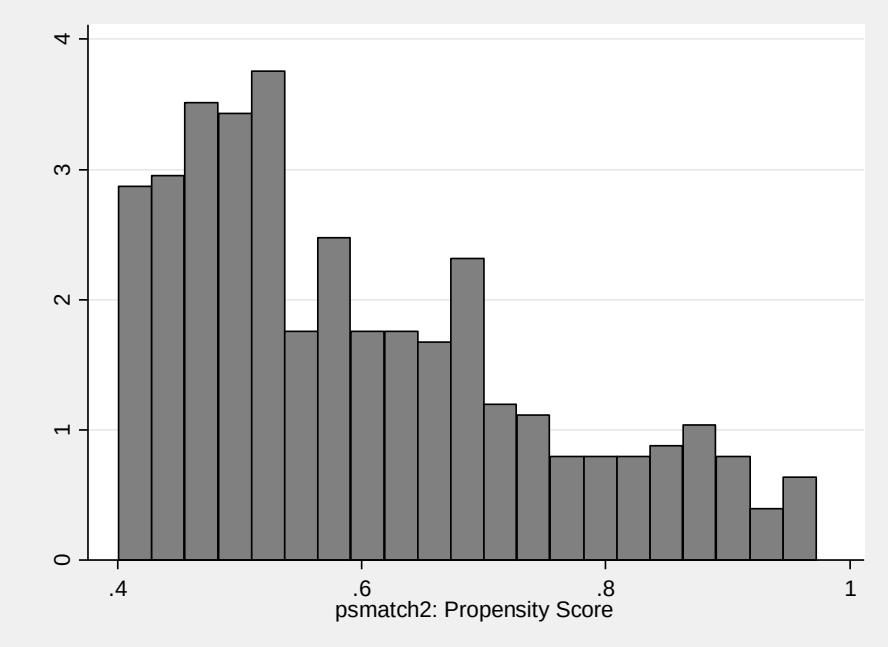


Figure 6-2 Density of Control Cases with Propensity Scores Greater than 0.4

After the propensity scores are estimated, individuals in the treatment group are then matched with individuals in the control group with similar propensity scores, or probability of participating in the program. There are a number of matching algorithms which can be employed. This study applies three matching algorithms that are commonly used.

The first one is nearest neighbor matching—for each treatment case, find the control case with the closest propensity score, —its "nearest neighbor". The control cases for which there are no treatment cases with a sufficiently similar score are discarded from the sample; the same is true for treatment cases for which there are no similar control cases. In standard matching without replacement, this control group member is then removed from the control reservoir and cannot be chosen again. However, in matching with replacement that was applied in this study, this control case is put back into the sample and can be used more than once to match other treated units as it is the best match. Each treated case is only used once, but the same control case

may be used over and over again if it is the closest match for many different treatment cases. The matched sample will consist of all the original treated cases, along with all the controls used as matches—and some of the controls will appear more than once in the matched data, since they can be used to match multiple treated cases. Matching with replacement tends to reduce bias relative to matching without replacement. Also matching without replacement cannot be performed when the control group is smaller than the treatment group. Once units are matched, the characteristics of the constructed treatment and comparison groups should not be significantly different; i.e., the matched units in the treatment and comparison groups should be statistically comparable.

Balance is generally tested using a t-test to compare the means of all covariates included in the propensity score in order to determine if the means are statistically similar in the treatment and comparison groups. If balance is not achieved, i.e., the means of the covariates are statistically different, a different matching option or specification should be used until the sample is sufficiently balanced. Table 6-3 shows the balance checks results from the nearest neighbor matching. The balance on covariates is examined by checking the standardized difference (STD) in the means between the matched control group and treatment group. It is not influenced by sample size and it allows for the comparison of the relative balance of variables measured in different units (Austin, 2011). Although there is no agreed criterion, a standard difference that is less than 0.1 has been taken to indicate a negligible difference in the mean or prevalence of a covariate between control and treatment groups (Normand et al., 2001). Based on the results, the STDs for all the variables are less than 0.1 after the matching, which means that there is no significant imbalance existing in any variables.

Table 6-3 Balance Checks of Propensity Score Matching

(Matching on have- salary sample using nearest neighbor matching)

		Pre-matched		Post-matched	
	Mean in Treated	**Mean in Untreated**	**Std. Diff.**	**Mean in Untreated**	**Std. Diff.**
Study migration	0.58	0.22	0.803	0.58	0.009
Age	23.06	22.97	0.089	23.1	-0.042
Female	0.3	0.47	-0.348	0.28	0.023
Minority	0.07	0.07	-0.01	0.08	-0.047
SES score	-0.35	-0.16	-0.207	-0.42	0.058
From rural area	0.56	0.43	0.253	0.55	0.022
From central or west areas	0.68	0.47	0.424	0.71	-0.056
Only child	0.3	0.38	-0.176	0.28	0.02
Family income	42488.05	50933.81	-0.22	40733.41	0.044
Student leader in high school	0.45	0.42	0.055	0.45	0.007
NCEE score	74.08	71.93	0.292	74.06	0.012
Humanities track in high school	0.11	0.22	-0.302	0.1	0.014
Arts or sports track in high school	0.03	0.05	-0.091	0.04	-0.059
Degree of preference towards one's major	2.67	2.61	0.071	2.67	-0.003
Had merit-based aid	0.34	0.33	0.006	0.34	-0.018
Had needs-based aid	0.3	0.25	0.123	0.32	-0.023
Had loan	0.35	0.29	0.119	0.36	-0.016
Percentage of work migration students	0.35	0.18	1.147	0.35	-0.028
Average course score in college	78.07	78.44	-0.06	77.9	0.017
Science or engineering major	0.75	0.57	0.395	0.74	0.051
Economics or management major	0.13	0.21	-0.212	0.14	-0.016
Have worked at college	0.84	0.87	-0.083	0.86	-0.034
Passed CET4	0.49	0.43	0.133	0.48	0.039
Passed CET6	0.37	0.34	0.061	0.37	-0.023
CCP member	0.35	0.33	0.048	0.36	-0.02
Student leader	0.19	0.22	-0.065	0.16	0.076
Institution located in municipalities	0.1	0.29	-0.499	0.08	0.038
Institution located in central or west areas	0.75	0.45	0.651	0.77	-0.031
985 institutions	0.18	0.07	0.335	0.18	0.003

211 institutions	0.33	0.3	0.062	0.34	0.004
Institution campus located in suburb	0.43	0.52	-0.185	0.43	-0.031
Independent institutions	0.02	0.01	0.073	0.01	0.057
Engineering-concentrated institutions	0.7	0.45	0.515	0.71	-0.016
Provincial GDP	5895.3	7685.01	-0.592	5718.98	0.051
Provincial population	4754.29	4688.24	0.024	5045.21	-0.103
Provincial size	20.05	20.97	-0.034	21.07	-0.034

The second algorithm is radius matching, or "caliper" matching. First a maximum propensity score radius —a caliper—is established, and then all control units within the given radius of a treatment unit are matched to that unit. Table 6-4 represents the balance checks results from the radius matching.

Table 6-4 Balance Checks of Propensity Score Matching

(Radius matching with a caliper of 0.05)

		Pre-matched		**Post-matched**	
	Mean in Treated	**Mean in Untreated**	**Std Diff.**	**Mean in Untreated**	**Std Diff.**
Study migration	0.58	0.22	0.803	0.57	0.019
Age	23.06	22.97	0.089	23.1	-0.037
Female	0.3	0.47	-0.348	0.3	-0.004
Minority	0.07	0.07	-0.01	0.07	-0.031
SES score	-0.35	-0.16	-0.207	-0.42	0.058
From rural area	0.56	0.43	0.253	0.58	-0.028
From central or west areas	0.68	0.47	0.424	0.71	-0.057
Only child	0.3	0.38	-0.176	0.28	0.031
Family income	42488.05	50933.81	-0.22	41067.84	0.034
Student leader in high school	0.45	0.42	0.055	0.46	-0.029
NCEE score	74.08	71.93	0.292	74.24	-0.013
Humanities track in high school	0.11	0.22	-0.302	0.1	0
Arts or sports track in high school	0.03	0.05	-0.091	0.04	-0.019

Degree of preference towards one's major	2.67	2.61	0.071	2.66	0.011
Had merit-based aid	0.34	0.33	0.006	0.33	0.014
Had needs-based aid	0.3	0.25	0.123	0.32	-0.034
Had loan	0.35	0.29	0.119	0.37	-0.041
Percentage of work migration students	0.35	0.18	1.147	0.35	-0.011
Average course score in college	78.07	78.44	-0.06	77.92	0.015
Science or engineering major	0.75	0.57	0.395	0.74	0.049
Economics or management major	0.13	0.21	-0.212	0.15	-0.045
Have worked at college	0.84	0.87	-0.083	0.86	-0.032
Passed CET4	0.49	0.43	0.133	0.49	0.026
Passed CET6	0.37	0.34	0.061	0.37	-0.012
CCP member	0.35	0.33	0.048	0.39	-0.085
Student leader	0.19	0.22	-0.065	0.19	0.004
Institution located in municipalities	0.1	0.29	-0.499	0.09	0.023
Institution located in central or west areas	0.75	0.45	0.651	0.76	-0.005
985 institutions	0.18	0.07	0.335	0.2	-0.072
211 institutions	0.33	0.3	0.062	0.33	0.018
Institution campus located in suburb	0.43	0.52	-0.185	0.43	-0.03
Independent institutions	0.02	0.01	0.073	0.01	0.043
Engineering-concentrated institutions	0.7	0.45	0.515	0.68	0.043
Provincial GDP	5895.3	7685.01	-0.592	5796.92	0.025
Provincial population	4754.29	4688.24	0.024	4904	-0.051
Provincial size	20.05	20.97	-0.034	20.66	-0.019

The third one is kernel matching: For each treated subject, a weighted average of the outcome of all control units is derived. The weights are based on the distance of the control group propensity score to that of the treated subject's, with the highest weight given to those with scores closest to the treated unit. Table 6-5 shows the results based on kernel matching.

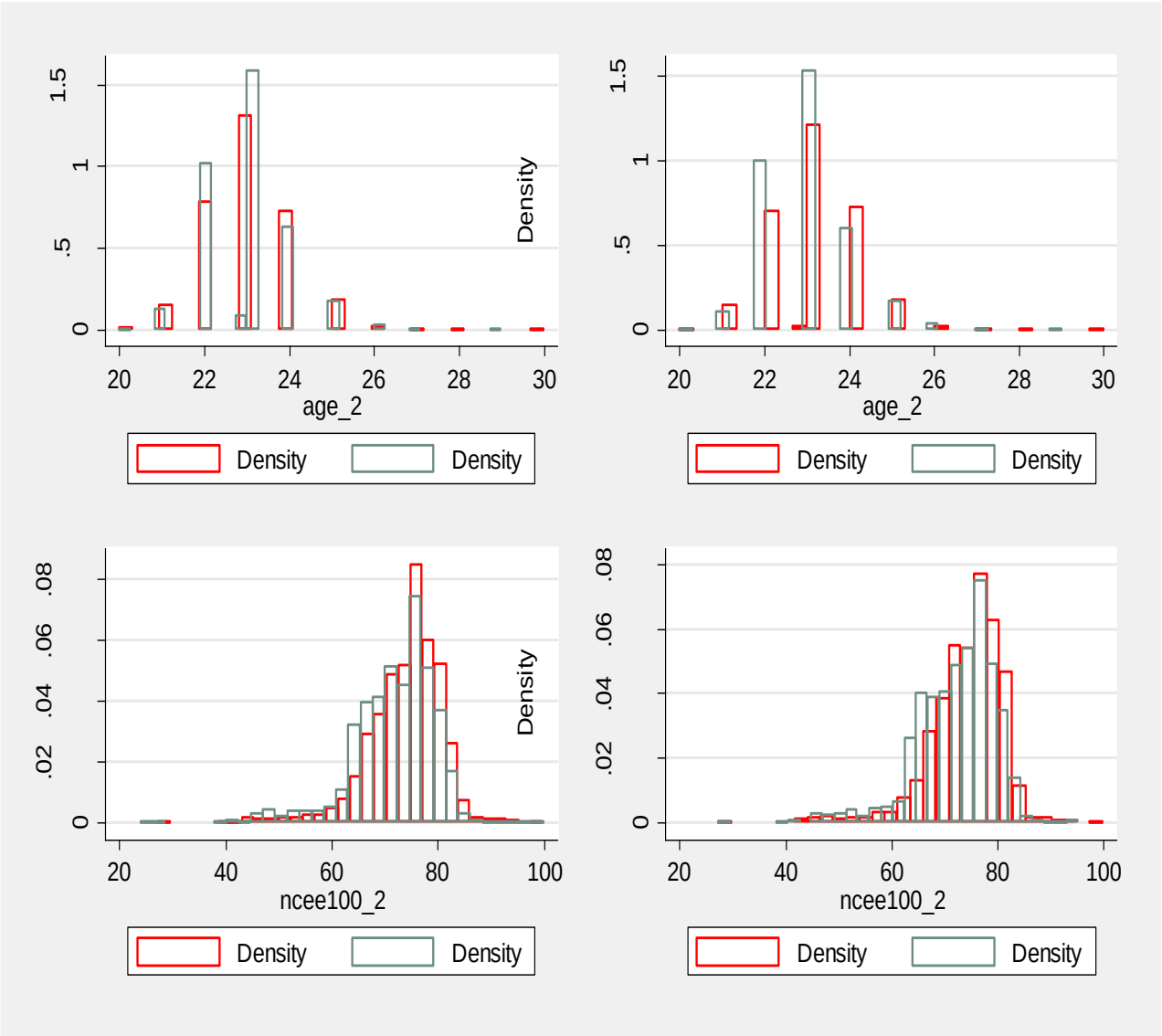
Table 6-5 Balance Checks of Propensity Score Matching

(Kernel matching)

		Pre-matched		Post-matched	
	Mean in Treated	**Mean in Untreated**	**Std. Diff.**	**Mean in Untreated**	**Std. Diff.**
Study migration	0.58	0.22	0.803	0.57	0.017
Age	23.06	22.97	0.089	23.09	-0.034
Female	0.3	0.47	-0.348	0.29	-0.003
Minority	0.07	0.07	-0.01	0.07	-0.03
SES score	-0.35	-0.16	-0.207	-0.42	0.059
From rural area	0.56	0.43	0.253	0.58	-0.031
From central or west areas	0.68	0.47	0.424	0.71	-0.058
Only child	0.3	0.38	-0.176	0.28	0.032
Family income	42488.05	50933.81	-0.22	41125.59	0.033
Student leader in high school	0.45	0.42	0.055	0.46	-0.032
NCEE score	74.08	71.93	0.292	74.25	-0.014
Humanities track in high school	0.11	0.22	-0.302	0.1	0.001
Arts or sports track in high school	0.03	0.05	-0.091	0.04	-0.02
Degree of preference towards one's major	2.67	2.61	0.071	2.66	0.011
Had merit-based aid	0.34	0.33	0.006	0.33	0.014
Had needs-based aid	0.3	0.25	0.123	0.32	-0.036
Had loan	0.35	0.29	0.119	0.37	-0.042
Percentage of work migration students	0.35	0.18	1.147	0.35	-0.013
Average course score in college	78.07	78.44	-0.06	77.92	0.014
Science or engineering major	0.75	0.57	0.395	0.74	0.048
Economics or management major	0.13	0.21	-0.212	0.15	-0.042
Have worked at college	0.84	0.87	-0.083	0.86	-0.032
Passed CET4	0.49	0.43	0.133	0.49	0.026
Passed CET6	0.37	0.34	0.061	0.37	-0.012
CCP member	0.35	0.33	0.048	0.39	-0.087
Student leader	0.19	0.22	-0.065	0.19	0.005
Institution located in municipalities	0.1	0.29	-0.499	0.09	0.023
Institution located in central or west areas	0.75	0.45	0.651	0.76	-0.008
985 institutions	0.18	0.07	0.335	0.2	-0.07

211 institutions	0.33	0.3	0.062	0.33	0.013
Institution campus located in suburb	0.43	0.52	-0.185	0.43	-0.027
Independent institutions	0.02	0.01	0.073	0.01	0.043
Engineering-concentrated institutions	0.7	0.45	0.515	0.68	0.041
Provincial GDP	5895.3	7685.01	-0.592	5795.41	0.025
Provincial population	4754.29	4688.24	0.024	4899.99	-0.049
Provincial size	20.05	20.97	-0.034	20.66	-0.019

The above three tables present the balance checks on the covariates after each matching process. Besides checking the STDs, the balance checks of the distribution of continuous variables are shown in the following figure. All three algorithms that are employed show that the balance has been successfully achieved.


0
.5
1
1.5
Density
20
22
24
26
28
30
age_2
Density
Density
0
.5
1
1.5
Density
20
22
24
26
28
30
age_2
Density
Density
0
.02
.04
.06
.08
Density
20
40
60
80
100
ncee100_2
Density
Density
0
.02
.04
.06
.08
Density
20
40
60
80
100
ncee100_2
Density
Density

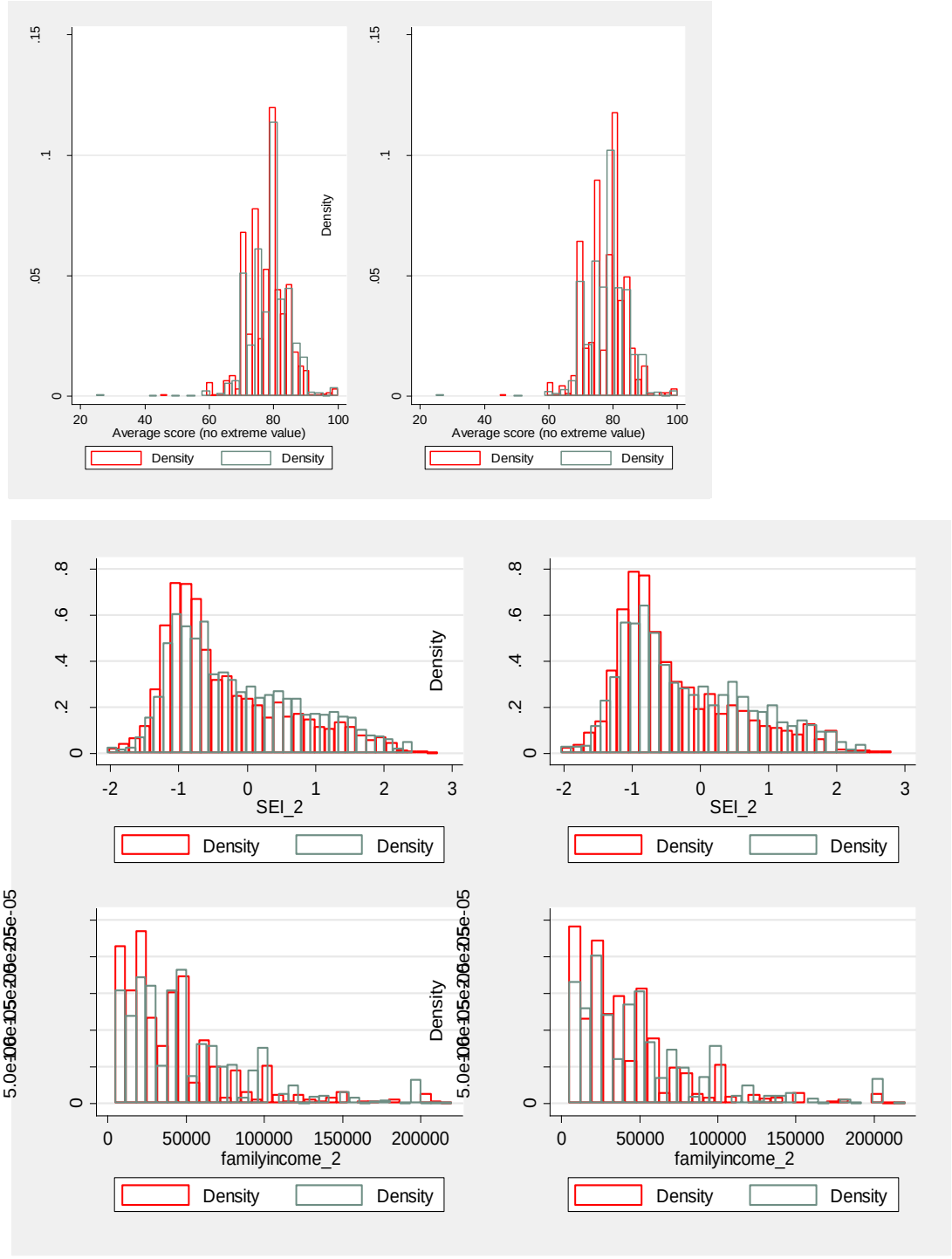


Figure 6-3 Density Distribution of Continuous Covariates in Treated and Untreated Groups
(Left: pre-matched; Right: post-matched) (Red: work migration=1; Teal: work migration=0)

Following the estimation of propensity scores, the implementation of a matching algorithm, and the achievement of balance, the intervention's impact (this is where we will see work migration's impact) is then estimated by regression (regression of outcomes on the treatment indicator and confounding covariates using weights to force the sample to represent matched groups). Table 6-6 presents the regression- adjusted matched estimates of the impact of work migration on starting monthly salary. The results from the three different algorithms are presented in the table. The covariates that were used in the previous OLS regression are also included in the model. The OLS estimates without sampling weights are also presented for comparison.

Table 6-6 PSM Estimates of the Impact of Work Migration on Starting Salary

	OLS w/o Weights	Nearest Neighbor Matching	Kernel Matching	Radius Caliper Matching
Work migration	0.0859***	0.0963***	0.0936***	0.0932***
	(0.0181)	(0.0206)	(0.0232)	(0.0231)
Age	0.0055	0.0013	0.0050	0.0048
	(0.0067)	(0.0069)	(0.0088)	(0.0088)
Female	-0.0821***	-0.0869***	-0.0630*	-0.0634*
	(0.0155)	(0.0174)	(0.025)	(0.025)
Minority	0.0265	0.0203	0.0091	0.0077
	(0.0263)	(0.025)	(0.0384)	(0.0385)
Only child	-0.0102	-0.0115	-0.007	-0.0070
	(0.0163)	(0.0186)	(0.0269)	(0.027)
From rural area	-0.025	-0.0215	-0.0625*	-0.0625*
	(0.0172)	(0.019)	(0.0294)	(0.0296)
SES score	0.0330***	0.0325**	0.0168	0.0171
	(0.0099)	(0.0113)	(0.0147)	(0.0148)
Science or engineering major	-0.0233	-0.0093	-0.0122	-0.0116
	(0.0242)	(0.027)	(0.0419)	(0.0419)
Econ or management major	-0.0835***	-0.0760**	-0.0835	-0.0841
	(0.0237)	(0.0276)	(0.0442)	(0.0444)
Student leader in high school	0.0272*	0.0273	0.0006	0.0013

	(0.013)	(0.0144)	(0.0216)	(0.0216)
Humanities track	-0.0555*	-0.0487	-0.0606	-0.0598
	(0.0225)	(0.0264)	(0.0404)	(0.0404)
Arts or sports track	0.0162	0.0137	-0.0802	-0.0781
	(0.0418)	(0.0522)	(0.0694)	(0.0694)
NCEE (rescaled to 1-100)	0.0058***	0.0056***	0.0042*	0.0042*
	(0.0013)	(0.0015)	(0.0020)	(0.0019)
Average course score	0.0008	0.0004	0.0047**	0.0047**
	(0.0013)	(0.0013)	(0.0016)	(0.0016)
Has a minor	-0.0060	-0.0196	-0.0339	-0.0353
	(0.0252)	(0.0299)	(0.0338)	(0.034)
Preference degree towards one's major	0.0252**	0.0254**	0.0198	0.0196
	(0.0083)	(0.0096)	(0.0128)	(0.0128)
Passed CET6	0.128***	0.133***	0.112***	0.112***
	(0.0215)	(0.0242)	(0.0301)	(0.0301)
Passed CET4	0.0724***	0.0679***	0.0672**	0.0671**
	(0.0185)	(0.0202)	(0.0251)	(0.0251)
Student leader	0.0336*	0.0345	0.0518	0.0512
	(0.0162)	(0.0186)	(0.0282)	(0.0283)
CCP member	0.0453**	0.0411*	0.0471	0.0477
	(0.0148)	(0.0163)	(0.0267)	(0.0269)
Has professional certificates	0.0115	0.0177	0.0301	0.0303
	(0.0127)	(0.0137)	(0.0206)	(0.0207)
Had merit-based aid	0.0191	0.0218	-0.0178	-0.017
	(0.0152)	(0.0166)	(0.0238)	(0.0239)
Had needs-based aid	-0.0225	-0.0197	-0.0112	-0.0118
	(0.0159)	(0.0159)	(0.0243)	(0.0243)
Had loan	-0.0195	-0.0203	-0.0246	-0.0251
	(0.0147)	(0.0165)	(0.0232)	(0.0233)
Comprehensive institutions	0.0700**	0.0617*	0.150***	0.150***
	(0.0243)	(0.0275)	(0.0449)	(0.045)
Engineering-concentrated	0.0492*	0.0244	0.068	0.0678
	(0.0231)	(0.0258)	(0.0368)	(0.0368)
985 institution	0.168***	0.164***	0.186***	0.187***
	(0.0259)	(0.0317)	(0.0468)	(0.0471)
211 institution	0.115***	0.114***	0.111***	0.111***
	(0.0167)	(0.0179)	(0.0265)	(0.0266)
Institution located in central or west areas	-0.0072	-0.0207	-0.0089	-0.0081
	(0.018)	(0.0192)	(0.024)	(0.0241)
Independent college	-0.0903	-0.0392	-0.0741	-0.0745
	(0.0523)	(0.0489)	(0.0546)	(0.0551)
Ever worked in college	-0.0473*	-0.0499*	-0.0362	-0.0361
	(0.0185)	(0.0209)	(0.0247)	(0.0249)

Constant	6.861***	7.056***	6.750***	6.758***
	(0.215)	(0.236)	(0.297)	(0.297)
N	3146	2548	2548	2548
R-square	0.2924	0.2917	0.3095	0.3104
R-square Adjusted	0.2694	0.2643	0.2828	0.2836

Notes: 1. Robust standard errors in parentheses;
2. Employment type dummies, provincial dummies and missing variables dummies are included in each model; however, the coefficients are not shown in the table;
3. * p<0.05, ** p<0.01, *** p<0.001.

From the results based on the three different types of matching process, it is clear that work migration has a significant positive impact on graduates' starting salary.

When comparing the OLS estimates without sampling weights and the PS-adjusted regression estimates, the results for work migration suggest that the magnitudes of the PSM estimates are larger than the OLS estimates. It means that the OLS estimates tend to be downward biased and the OLS regression underestimated the positive impact of the work migration on the starting salary. Work migration has a statistically significant positive impact on graduates' starting salaries. Explicitly, the estimated coefficients are 0.0963, 0.0936, and 0.0932 for PS-adjusted regressions based on nearest neighbor matching, kernel matching, and radius caliper matching schemes.

Similar to the OLS results, the PS-adjusted regression estimates show that being female has a significant negative impact on starting salary. Having a higher NCEE score, having passed the CET4 and/or CET6, having studies at a comprehensive institution, or at a 985 or 211 institution, all have a significant positive impact on a graduate's starting salary. There is slight difference between the results from nearest neighbor matching and the other two matching algorithms. From the results of kernel and radius caliper matching algorithms, the variable, from rural area, has a statistically significant negative impact on student's starting salary; average course score in

college has a statistically significant positive impact on starting salary. For results from nearest neighbor matching, the variables SES score, the preference degree towards one's degree, and being a CCP member all have a statistically significant positive impact on student's starting salary while the variables being an economics or management major and have worked in college are the ones that have a statistically significant negative impact on starting salary.

In summary, the PSM estimates indicate a statistically significant positive impact of work migration on the graduate's starting monthly salary. The finding is consistent with the basic unweighted OLS results. The difference in the magnitude between the PSM estimates and the OLS estimates without sampling weights suggests that the basic model estimates tend to be downward biased. This finding may indicate that selection into work migration might be negatively related to a student's innate ability. As discussed before, PSM still relies on a very strong assumption, a conditional-independence assumption (CIA), which requires that the common variables that affect treatment assignment and treatment-specific outcomes be observable. Because the propensity score is constructed with observables and the CIA might not be satisfied, the propensity score method does not solve the omitted variable bias and therefore cannot fully address the endogeneity problem. Therefore, the PSM estimates cannot be interpreted as the causal impact of work migration on starting salary.

Section 6.4 Impact of Work Migration on Starting Salary, Method of Instrumental Variable

The Instrumental Variable (IV) method uses a proxy variable approach to mitigate or even eliminate the bias posed by problems such as endogeneity, measurement error and omitted variables (Wooldridge, 2002). The IV approach leaves the unobservable factor in the residual of

the structural equation, instead of modifying the set of moment conditions used to estimate the parameters (Soderbom, 2009).

Omitted variable bias is the most common illustration of what economists refer to as endogeneity. In this study, there might be some unobserved endogenous variables of initial salary (e.g., unobserved ability) are not included in the model. In other words, a college graduate's migration behavior is possibly correlated with certain unobservable student characteristics such as individual innate ability. Omission of such variables, particularly when they are also correlated with included explanatory variables will cause bias in identifying the link between initial salary and the included variables.

In this section, a two-stage least square model (2SLS) is estimated to try to more accurately examine the impact of work migration on graduates' starting salary and provide consistent estimates of the parameters. The procedure is as follows: first regress the endogenous variable on the instrument(s) using OLS and calculate the predicted values of the endogenous variable. Then use the predicted value from the first regression as an explanatory variable in the starting-salary equation, and estimate the equation using OLS. The resulting estimate of the coefficient on the predicted variable is the IV estimate. A valid instrument isolates a part of the independent variable that is uncorrelated with the error term, and that part can be used to estimate the effect of a change in independent variable on the outcome variable. We can interpret this in terms of "purging" the endogenous variable of the correlation with the residual.

It might be better to have a moderately over identified model (more instruments than strictly needed), because the additional instruments can be used to increase the precision of the estimates (Soderbom, 2009). By assumption all the instruments produce exogenous variation in the

predicted variable and the OLS estimation in the first stage ensures there is as much such variation as possible. With fewer instruments there would be less exogenous variation in this predicted variable, hence such estimators would not be efficient. In this study, two instrumental variables were proposed for work migration: "percentage of students migrating to work in each institution" and the "employment outflow rate of the province where the institution is located". Table 6-7 shows the results for two different regressions: the first includes only one instrument percentage of students migrating to work in each institution and the second includes both of the two instruments in the model.

The IV method is based on two main assumptions. Relevance: IVs are sufficiently correlated with the included endogenous regressors, conditional on all exogenous variables in the model. Validity or exclusion restrictions: the excluded instruments are distributed independently of the error and they affect the dependent variable only indirectly, through their correlations with the included endogenous variables. The first relevance condition can be tested by computing the *t-statistic* associated in the first-stage regression by checking the correlation between the instruments and the probability of treatment. Tests of overidentifying restrictions address the second assumption, a rejection of the null hypothesis may be indicative that the exclusion restrictions for these instruments may be inappropriate. In other words, some of the instruments may have been incorrectly excluded from the regression model's specification.

Instruments may be satisfactorily exogenous, but only weakly correlated with the endogenous regressors. Since Staiger and Stock (1997) formalized the definition of weak instruments, many researchers have concluded the "rule of thumb" that if the first-stage *F* statistic exceeds 10, their instruments are sufficiently strong. However, the weak instruments problem can arise in a large sample even when the first-stage *t*- and *F*-tests are significant at

conventional levels. In the worst case, the IV estimator bias is the same as that of the OLS, the IV therefore is inconsistent, and nothing is gained by instrumenting. With sufficiently weak instruments, the asymptotic identification status of the equation is called into question. The .first command option of ivreg2 in Stata presents several useful diagnostic statistics that can be used to critically evaluate the strength of instruments. For example, the test proposed by Stock and Yogo (2005) provides useful rules for evaluating the weakness of instruments. Stock-Yogo test is based on the performance of the Wald test statistic for the endogenous regressors. Under weak identification, the test rejects the null hypothesis too often. The test statistic is based on the rejection rate r which is tolerable to the researcher if the true rejection rate is 5%. The endogenous regressors' tabulated values consider various values for r. To be able to reject the null hypothesis that the size of the test is unacceptably large (versus 5%), the Cragg–Donald F statistic must exceed the tabulated critical value.

Table 6-7 reports the instrumental variable estimates of the 2SLS model using the graduates' starting monthly salary as the outcome variable. The first-stage coefficients, the Kleibergen-Paap rk LM statistics, the Kleibergen-Paap Wald rk F statistics, the Cragg-Donald Wald F statistics, the Anderson-Rubin Wald weak-instrument robust test, the Wu-Hausman test and the Sargan statistic are reported in the lower half of the table. In order to relate the validity condition to economic theory for the analysis to be more convincing, some falsification tests were done to test the validity of the instrumental variables.

Table 6-7 IV Estimates of the Impact of Work Migration on Starting Salary

(Dependent variable: Monthly starting salary in log form)

	(1) One IV	(2) Two IVs
Work Migration	0.203**	0.158*
	(0.068)	(0.0621)
Age	0.00412	0.00469
	(0.00623)	(0.00619)
Female	-0.0793***	-0.0802***
	(0.0157)	(0.0156)
Minority	0.029	0.028
	(0.0236)	(0.0235)
Only child	-0.01	-0.00996
	(0.0166)	(0.0165)
From rural area	-0.0273	-0.0265
	(0.0174)	(0.0173)
SES score	0.0326***	0.0327**
	(0.0101)	(0.0101)
Student leader in high school	0.0266*	0.0268*
	(0.0129)	(0.0128)
Humanities track in high school	-0.0522*	-0.0536*
	(0.0233)	(0.0232)
Arts or sports track in high school	0.0161	0.016
	(0.0472)	(0.047)
NCEE score	0.00558***	0.00565***
	(0.00134)	(0.00134)
Average course score	0.000869	0.000855
	(0.00124)	(0.00123)
Science or engineering major	-0.0229	-0.023
	(0.024)	(0.0236)
Economics or management major	-0.0804***	-0.0815***
	(0.024)	(0.0239)
Has a minor	-0.00378	-0.00453
	(0.026)	(0.0264)
Preference degree towards one's major	0.0229**	0.0238**
	(0.00852)	(0.00844)
Pass CET4	0.0649***	0.0675***
	(0.0180)	(0.0179)
Pass CET6	0.124***	0.125***
	(0.0214)	(0.0213)
Student leader in college	0.0336*	0.0335*

	(0.0166)	(0.0165)
CCP member	0.0483**	0.0471**
	(0.0151)	(0.015)
Has professional certificates	0.00973	0.0103
	(0.0125)	(0.0125)
Has worked in college	-0.0425*	-0.0444*
	(0.0180)	(0.0178)
Had merit-based aid	0.0193	0.0191
	(0.015)	(0.0149)
Had needs-based aid	-0.0227	-0.0227
	(0.0146)	(0.0145)
Had loan	-0.0211	-0.0204
	(0.0146)	(0.0145)
Comprehensive institutions	0.0844***	0.0788**
	(0.0248)	(0.0247)
Engineering-concentrated institutions	0.0298	0.0372
	(0.0260)	(0.0253)
985 institution	0.136***	0.148***
	(0.0317)	(0.0309)
211 institution	0.112***	0.113***
	(0.0164)	(0.0164)
Independent college	-0.0902+	-0.0903
	(0.0471)	(0.0469)
Institution located in central or west area	-0.0468	-0.0314
	(0.0267)	(0.0254)
N	3146	3146
R-squared	0. 283	0.29

IV First-stage Regression Outputs

Endogenous variable		
Percentage of students migrating to work in each institution	1.08***	0.927***
	(0.086)	(0.0944)
Employment outflow rate		0.004***
		(0.001)
N	3146	3146
R-squared	0.5667	0.5715
Kleibergen-Paap rk LM stat (under identified test)	150.02***	175.85***
Kleibergen-Paap Wald rk F stat (weak IV)	157.24***	92.97***
Cragg-Donald Wald F stat (weak IV)	220.49***	128.58***
Anderson-Rubin Wald weak IV robust test Chi-sq	8.73**	9.64**
Wu-Hausman F-stat p-value	0.07	0.04
Sargan stat		3.241

Notes: 1. Sampling weights are applied and robust standard errors are in parentheses
2. Industry, employer type, province of workplace, and missing dummies are included
3. + p<0.10, * p<0.05, ** p<0.01, *** p<0.001

Based on the regression results with one instrumental variable shown in the Column 1 of Table 6-7, work migration has a significantly positive effect on fresh college graduates' starting salaries is at the 1% level. When all the other covariates are held constant, college graduates who would migrate to work enjoy a 20.3% starting salary premium over those who would not migrate to work in China.

The results also show that students' individual characteristics, family background, academic achievement, college experience and institutional attributes all influence their early post-college labor market outcomes. Specifically, being a female student is associated with a 7.93% decrease in starting monthly salary. College graduate who was a student leader in high school will have 2.9% higher starting monthly salary. Similarly, being on the humanities track in high school is associated with a 5.22% decrease in starting salary. A one-unit increase in a student's SES score is associated with 3.26% increase in starting monthly salary. For a ten-unit increase in NCEE score, we would expect to see about a 5.58% increase in starting salary.

In terms of college experience and institution attributes, majoring in economics or management is associated with an 8.04% decrease in stating monthly salary. A one- unit increase in the preference degree towards one's major is associated with a 2.29% increase in a graduate's starting salary. Compared with students who didn't pass CET4, students who passed the CET4 have a 6.49% higher starting monthly salary. Students who passed CET6 have a 12.4% higher starting salary. Being a student leader in college is associated with a 3.36% increase in a new graduate's starting salary. Compared with students who are not CCP members, students who

joined CCP have a 4.83% higher starting salary. For college work experience, having worked in college is associated with a 4.25% decrease in a graduate's starting salary.

If a student is from a comprehensive institution, there will be an 8.44% increase in salary compared to those from institutions with other concentrations other than engineering. Compared with students from non-key higher education institutions, students graduating from a 985 institution have a 13.6% higher starting salary. Similarly, compared with students from non-key higher education institutions, students studying in a 211 institution have an 11.2% higher starting salary. In addition, graduating from an independent institution is associated with a 9.02% decrease in new graduate's starting salary.

Similar results were found based on the regression results with two instrumental variables shown in the Column 2 of Table 6-7. Work migration has a significantly positive effect on new college graduates' starting salaries at the 5% level. College graduates who would migrate to work enjoy a 15.8% starting salary premium over those who would not migrate to work in China.

Students' individual characteristics, family background, academic achievement, college experience and institution attributes all impact on a student's early post-college labor market outcomes. In more details, female students have an 8.02% lower starting monthly salary than their male counterparts. Being a student leader in high school is associated with a 2.68% increase in starting monthly salary. Similarly, being on the humanities track in high school is associated with a 5.36% decrease in starting salary. A one-unit increase in the student's SES score is associated with 3.27% increase in starting monthly salary. For a ten-unit increase in NCEE score, we would expect to see about a 5.65% increase in starting salary.

In terms of college experience and institution attributes, having a major in economics and management is associated with an 8.15% decrease in stating monthly salary. A one-unit increase

in the preference degree towards one's major is associated with a 2.38% increase in the graduate's starting salary. Having Passed the CET4 and CET6 is associated with a 6.75% and 12.5% increase respectively in starting monthly salary. Being student leader in college is associated with a 3.35% increase in new graduate's starting salary. CCP members have a 4.71% higher starting salary than those graduates who are not CCP members. For college work experience, having worked in college is associated with a 4.44% decrease in a graduate's starting salary.

Students graduating from a comprehensive institution will have a 7.88% higher salary compared with those graduating from institutions with other concentrations. Graduates studying in a 985 institution will have a 14.8% higher starting salary compared with those in non-key higher institutions. Similarly, studying in a 211 institution is associated with an 11.3% increase in new graduate's starting salary.

Consistent with the PSM findings, the magnitude of the coefficient estimate for work migration in IV regression is larger than the OLS estimate. Although IV strategy is useful in constructing proper comparison group, when interpreting the results from IV regression, we need to keep in mind that without constant treatment effect assumption[13] we can only identify average effects for subpopulations that are induced by the instrument to change the value of the endogenous regressors. We refer to such subpopulations as compliers, and to the average treatment effect as the local average treatment effect (LATE). These complier subpopulations are not necessarily the subpopulations that we are most interested in and researchers should be careful to generalize the average effects for other subpopulations (Imbens & Wooldridge, 2007;

[13] The constant causal effects assumption holds that a causal effect will be the same across units or across time within a unit. In other words, the magnitude of some causal effect does not depend on the characteristics of a unit or of the period in which it receives the treatment. Rather, every unit that receives a treatment will react in the same manner.

Loewen, 2010). In this dissertation study, the reason for the larger IV estimates might be that it only reveals the LATE on students whose work migration status is influenced by the instrumental variables, i.e. the percentage of students migrating to work in each institution and employment outflow rate. These students may be different from other students and they are more susceptible to institution/provincial climate, as they do not have a clear incentive of work migration or not. The larger IV estimate might reflect a heterogeneous effect of work migration by students' motivation and ability.

However, as it is not possible to identify the affected sample with available data, it is difficult to decide whether the above speculations about the sample are correct or not. Therefore it is not clear whether the IV estimates of the impacts of work migration on labor market outcomes are generalizable to all students who chose to migrate to work.

Validity of the instrumental variable

There are some key assumptions underlying IV analysis: the instrument itself is randomly assigned, it has a clear impact on actual treatment probability, but instrument has no effect on outcomes except through the treatment of interest (no direct effect, no treatments associate with instrument other than the one of interest).

The first assumption requires that the instrument has a clear impact on actual treatment probability. The first-stage regression shows that the two instrumental variables percentage of students migrating to work in each institution and employment outflow rate have a significant effect on work migration. As shown in the above Table 6-7, the first-stage coefficients for the IVs are statistically significant in both models. One percentage point increase in the instrument variable percentage of students migrating to work in each institution is associated with an

increase of 1.08 points in the probability of work migration. In the two-IV model, one percentage point increase in the percentage of students migrating to work in each institution leads to an increase of 0.93 points in the probability of work migration. One percentage point increase in the employment outflow rate leads to an increase of a 0.5 percentage points in the probability of work migration.

The Kleibergen-Paap Wald F statistics and Cragg-Donald Wald F-statistics in both models are all much greater than 10, indicating that a weak IV is not a problem here. In addition, the Anderson-Rubin weak instrument robust test (tests of joint significance of endogenous regressors in the second equation) show that the coefficient of work migration is statistically significantly different from zero, indicating that there is a positive significant impact of work migration on starting salary.

Because there are two instrumental variables included in the second model, we need to test whether the excluded instruments are appropriately independent of the error process and evaluate the validity of the instruments. Under the assumption of i.i.d. errors[14], the command ivreg2 routinely produces the Sargan test of overidentifying restrictions, which estimates the structural equation by 2SLS and obtains the residuals $\widehat{\varepsilon_i}$ and then regresses the residuals on all exogenous variables and obtains the R-squared, R_1^2. Under the null hypothesis that all instrument variables are uncorrelated with ε, the test has a $\chi^2(r)$ distribution where r is the number of overidentifying restrictions (the number of IVs from outside the model minus the total number of endogenous explanatory variables). If the test statistic exceeds the critical value we reject the null hypothesis and conclude that at least some of the IVs are not exogenous. In this study it shows that the

[14] In probability theory and statistics, a sequence of random variables is independent and identically distributed (i.i.d.) if each random variable has the same probability distribution as the others and all are mutually independent.

Sargan statistic is not significant so that it cannot reject the null hypothesis "overidentifying restrictions are valid".

The Wu-Hausman F test examines whether the correlation between the residuals of the logit equation and the reduced form equation is statistically significantly different from zero. Rejection of the null hypothesis indicates that there is an endogenous problem in the OLS estimation. The results show that the Wu-Hausman F test marginally rejects the null hypothesis of exogeneity in the first model with one instrument, and it rejects the null hypothesis in the second model with two instruments, suggesting that endogeneity is an issue for the OLS estimation that needs to be addressed. Previous analyses and discussions also provide some evidence of the existence of the endogeneity problem, and the PSM analysis in the previous section shows that the OLS estimates tend to be downward biased. Therefore the IV estimates are preferable to the OLS estimates.

The second assumption requires that the IVs do not have any direct impact on outcomes and are not associated with other treatments except for work migration. The only channel that instrument variable(s) can affect the outcome variable is through the treatment (Angrist & Pischke, 2009). The main concern is that the percentage of students migrating to work in one's institution may be correlated with potential early labor market outcomes through ways other than work migration, after controlling for all other covariates. For example, it is possible that more students have migrated to work in other provinces and consequently more employers get to know the institutions well. Therefore, such students might be offered higher starting salary than students from other institutions. However, this reputational impact is more likely to be the result of students who were already in the labor market from previous cohorts, rather than students who are still enrolled in college and searching for jobs. Another example might be the percentage of

graduates migrating to work in an institution may be higher because there is a more intensive recruitment effort made in that institution by employers, which is then also correlated with higher wage offers made to students in that institution. Similarly, the instrumental variable employment outflow rate does not have a relationship with starting salary other than through work migration.

Section 6.5 Summary of Empirical Findings

This study used the intention-to-work subsample from the CSLM 2011 survey data to explore the second research question—What is the impact of work migration on a new college graduate's starting salary? When exploring work migration's effects on a graduate's starting salary, in addition to the OLS method, the study also incorporated alternative identification strategies (instrumental variable method and PSM method) to address the issues from the potential endogenous treatment variable, work migration.

The weighted OLS analysis results suggest that holding all the other covariates constant, new graduates who decide to migrate to work enjoy a 9.9% starting monthly salary premium over those graduates who do not migrate in China. When using the unweighted sample, the OLS regression results show that the coefficient for work migration is 0.0859.

Three different PSM schemes were used: nearest neighbor matching, kernel matching, and radius caliper matching. The regression results from the propensity score-adjusted regression also showed that work migration has a significant positive impact on a student's starting salary. Work migration has a statistically significant positive impact on graduates' starting salaries. Explicitly, the estimated coefficients are 0.0963, 0.0936, and 0.0932 for PS-adjusted regressions based on nearest neighbor matching, kernel matching, and radius caliper matching schemes.

When comparing the OLS estimate without sampling weights and the propensity score-adjusted regression estimate, the results suggest the magnitude of the PSM estimate is larger than the unweighted OLS estimate, which means that the impact of work migration might be underestimated by OLS.

Consistent with the PSM findings, the magnitude of the coefficient estimate for work migration in IV regression is larger than the OLS estimate. Specifically, for the regression using one instrumental variable, work migration has a significantly positive effect on new college graduates' starting salaries at the 1% level. Graduates who would migrate to work enjoy a 20.3% salary premium over those who did not migrate. For the regression using two instrumental variables, work migration also has a significantly positive effect on new college graduates' starting salaries at the 5% level. College graduates who migrated to work will have a 15.8% higher starting salary compared to those who would not migrate to work.

Previous literature has mixed results about the impact of work migration on students' post-college labor market performance. The findings of this study about the significant positive impact of work migration on graduates' starting salary are consistent with some studies (Yue & Zhou, 2005; Ma, 2010; Yue, 2011); while others (Li, Zhao, & Guo, 2010) found no influence of work migration on students' labor market outcomes. Among these previous studies, Ma (2010) conducted a 2SLS regression to examine the relationship between college graduates' migration and their initial salary. She concluded that college graduates' work migration did help increase their initial salary. The coefficient associated with work migration reported in her study is 8.1%. Compared with the larger OLS estimate, she indicated that an endogeneity problem exists and the OLS coefficient overestimates the impact of migration on a graduate's starting salary. In Ma's (2010) OLS and two-stage least square regressions, only a few factors (gender, household

income, father's year of education, GDP per capita) were included in the model and there were no variables for institution characteristics included in the regression at all, which might lead to questionable conclusions.

This study uses the most recent representative sample of the college graduates survey in China (CSLM 2011), which means its conclusions are more easily applied to the current related population. In addition, the data include comprehensive information on the students, their families, the institutions and jobs; all such related information was included in the regression to explain the variance in a graduate's starting salary more accurately. This study incorporated the IV technique to examine the research question and it is the first study to employ the PSM identification strategy to address issues with the potential endogenous treatment variable, migration.

OLS, IV and PSM methodologies all have advantages and disadvantages. Different from OLS models, the PSM strategy does not have to specify the multi-dimensional relationship between explanatory variables and the outcome variable, but it uses a one-number summary of them to control for predictors. Propensity score theory says that rather than controlling for (stratifying on, regressing on, matching on) all the variables in X, it is sufficient to control for just the propensity score, e(x), which is just a one-number summary of X.

The main difference between the PSM and IV identification strategies is that propensity scoring matching method employs observable measures to construct a weight based on selection while IV method relies on an instrument variable made from unmeasured or unobserved factors. An advantage of IV is that it accounts for unmeasured factors correlated with the outcome. This is especially helpful to analyze the non-experimental data sets that were not created for the

purpose of the research question. Weakness of IV is that it might be challenging to find the instrument and it could be difficult to validate.

Based on previous research results from the literature review and findings from this study, it is obvious that work migration, student characteristics, family background, academic performance, and institution attributes all have an impact on students' early labor market outcomes. The following table is a summary table that compares the findings from previous literature with this study on the factors that affect fresh graduates' starting salaries.

Consistent with previous literature (Qing & Zeng, 2009; Du &Yue, 2010; Guo, Tsang, & Ding, 2010; Lai, 2012), the results show that female graduates are in a disadvantageous position in China's labor market. Being a female student is associated with about an 8% decrease in starting monthly salary. Yue (2011) found that being an only child has a statistically significant positive impact on a graduate's starting salary. However, this study showed that being the only child in a family has no significant impact on starting salary. In terms of family background, some studies (Yue & Zhou, 2005; Du & Yue, 2010; Lai, Meng, & Su, 2012) find that SES has a statistically significant positive impact on students' labor market outcomes; while other studies (Ren, Guo, & Pan, 2013) find no impact. Other family background information including age, race, and whether from a rural area are not significantly associated with starting salary. Though the magnitude is small, students' NCEE scores also have a statistically significantly positive impact on starting monthly salary, which is consistent with previous findings (Guo, Tsang, & Ding, 2010; Shi, et al, 2012).

Whether student was on humanities track in high school has a statistically significant negative impact on monthly starting salary. A measure of student innate ability—student leader in high school has a significant positive impact on new graduate's starting salary.

In terms of students' academic performance, there is mixed evidence from previous literature in China. While some studies (Yue & Zhou, 2005; Du & Yue, 2010; Ren, Guo, & Pan, 2013) find that academic performance has a significantly positive impact on monthly starting salary, some other studies (Guo, Tsang, & Ding, 2010; Lai, Meng & Su, 2012) find it has a negative impact; Huang (2007) finds no statistically significant impact. In this study, the results show that the average course score in college has no statistically significant association with the graduate's starting salary. As for the impact of passing the English proficiency tests CET4 and CET6 on graduates' starting salary, the findings from this study are consistent with the previous research (Du & Yue, 2010; Guo, Tsang & Ding, 2010; Li, Meng, & Shi, 2012; Lai, Meng, & Su, 2012). Passing the CET4 or CET6 has a statistically significant positive impact on a graduate's starting salary.

With regards to college experience, the students' major significantly influenced their labor market outcomes. It was found that students with an economics or management major tended to have a lower starting salary than humanities students. This phenomenon might be attributed to the much larger number of new graduates who majored in economics and management than were actually needed in recent years. In addition, students with a more positive attitude towards their major had a statistically significant higher starting salary. As for other college activities, whether the student is a CCP member, and whether they were a student leader in college are important factors in influencing students' labor market outcomes and are both significantly positively associated with starting salary. This is consistent with findings from previous studies (Huang, 2007; Du & Yue, 2010; Yue, 2011; Lai et al., 2012; Li, et al., 2012; Shi, et al, 2012; Ren, et al., 2013).

With regards to institution characteristics, both the institution's academic ranking and concentration had a significant positive impact on graduates' labor market outcomes. The starting monthly salaries of graduates from 985 institutions were about 15% higher than for graduates from non-key institutions. The starting monthly salaries for graduates from 211 institutions were about 12% higher than for graduates from non-key institutions. On the whole, these findings are consistent with previous studies (Du & Yue, 2010; Guo, Tsang, & Ding, 2010; Yue, 2011; Li, Zhao, & Guo, 2010).

Table 6-8 Comparison of Studies on a Range of Variables' Impact on Starting Salary

Variables	Previous studies	Impact	This study
Work migration	Yue & Zhou (2005)	Positive	Positive
	Li, Zhao & Guo (2010)	No	
	Ma (2010)	Positive	
	Yue (2011)	Positive	
Being female	Du & Yue (2010)	Negative	Negative
	Guo, Tsang, & Ding (2010)	Negative	
	Lai (2012)	Negative	
	Qing & Zeng (2009)	Negative	
Being only child	Yue (2011)	Positive	No
From rural area	Du & Yue (2010)	No	No
SES score	Yue & Zhou (2005)	Positive	Positive
	Du & Yue (2010)	Positive	
	Lai, Meng, & Su (2012)	Positive	
	Ren, Guo, & Pan (2013)	No	
NCEE score	Guo, Tsang, & Ding (2010)	Positive	Positive
	Shi, et al (2012)	Positive	Small magnitude
Academic performance	Ren, Guo, & Pan (2013)	Positive	No
	Du & Yue (2010)	Positive	

	Guo, Tsang, & Ding (2010)	Negative	
	Lai, Meng, & Su (2012)	Negative	
	Huang (2007)	No	
	Yue & Zhou (2005)	Positive	
CET4	Li, Meng, & Shi (2012)	Positive	Positive
	Lai, Meng, & Su (2012)		
	Du& Yue (2010)		
CET6	Guo, Tsang, & Ding (2010)	Positive	Positive
	Lai, Meng, & Su (2012)	Positive	
	Du & Yue (2010)	Positive	
CCP member	Yue (2011)	Positive	Positive
	Shi, et al (2012)	Positive	
211 institutions	Yue (2011)	Positive	Positive
	Du & Yue (2010)	Positive	
	Guo, Tsang, & Ding (2010)	Positive	
	Li, Zhao, & Guo (2010)	Positive	

In summary, the empirical analysis presented in this chapter suggests a statistically significant positive impact of work migration on fresh graduates' early post-college labor market performance--starting salary.

CHAPTER SEVEN:
CONCLUSIONS

China has accelerated the pace of higher education expansion since 1999 and entered the stage of mass higher education. Along with this achievement, an increasing rate of college graduates' unemployment has caused concerns in society, and such unemployment closely relates to economic development, education policy-making, and reforms in the economy as well as in higher education (Bai, 2006).

This dissertation examines the determinants of college graduates' work migration decisions in China and migration's impact on college graduates' early labor market performance, using nationally representative CSLM 2011 survey dataset. This study explicitly explores two key research questions by employing different quantitative techniques and models:

RQ1: What are the determinants of college graduates' migration decision in China?

RQ2: What is the impact of migration on college graduates' starting salaries?

In this chapter, the key findings are summarized in section 7.1, followed by discussion on this study's significance in section 7.2. Section 7.3 presents the study's limitations and suggestions for future research. Section 7.4 explores policy implications based on the key findings.

Section 7.1 Summary of Findings

This section presents a summary of key findings from this dissertation study. Section 7.1.1 describes the current college graduates' migration situation based on the national dataset and answers the first research question. Section 7.1.2 presents the quantitative findings on the impact of work migration on college graduates' starting salaries to answer the second research question.

7.1.1 Quantitative findings: determinants of college graduates' migration decisions in China

As shown in the weighted results, within the sampled 2007 college student cohort, 27.51% of the students who had offer(s) at the time of the survey reported that they would migrate to work for their new job after graduation (5.36% planned to go back to their hometown for work, 22.15% chose to work in a province that is neither their hometown nor where they went to college) while 36.29% said they would be working in the province where their higher education institution is located.

In order to examine the first research question "What are the determinants of college graduates' work migration decisions in China," we first used probit model when considering the two options "migrate" and "do not migrate". The multinomial model was then estimated to deal with the three choices of "college stayer", "return migrant" and "repeat migrant".

The results reveal that the following variables have a significant positive impact on college graduates' work migration: study migration, science or engineering major, student leader in high school, passed CET4, passed CET6, engineering-concentrated institution, from 985 institutions, and from 211 institutions. College graduates who possess the above individual characteristics or are from the above institutions are more likely to migrate to work. This study also finds that the following variables have a significant negative impact on work migration: female, provincial GDP per capita, provincial population, provincial area size, and provincial ECI score.

7.1.2 Quantitative findings: the impact of work migration on college graduates' starting salaries

This study used the intention-to-work subsample from the CSLM 2011 survey data to explore the second research question—What is the impact of work migration on new college

graduates' starting salaries in China? When exploring work migration's effects on a graduate's starting salary, in addition to the OLS method, the study also incorporated alternative identification strategies (instrumental variable method and PSM method) to address the issues from the potential endogenous treatment variable, work migration.

The unweighted OLS regression analysis reveals that when holding all the other covariates constant, new graduates who decide to migrate for work enjoy an 8.59% starting monthly salary premium over those graduates who do not do so. The weighted OLS regression analysis results suggest that new graduates who decide to migrate for work enjoy a 9.9% starting monthly salary premium over those who do not migrate in China.

Three different PSM schemes were used to conduct PS-adjusted regressions. The regression results showed that work migration has a statistically significant positive impact on graduates' starting salaries. Explicitly, the estimated coefficients are 0.0963, 0.0936, and 0.0932 for PS-adjusted regressions based on nearest neighbor matching, kernel matching, and radius caliper matching schemes. When comparing the OLS estimate without sampling weights and the propensity score-adjusted regression estimate, the results suggest the magnitude of the PSM estimate is larger than the unweighted OLS estimate.

Consistent with the PSM findings, the magnitude of the coefficient estimate for work migration in IV regression is larger than the OLS estimate. Specifically, for the regression using one instrumental variable, work migration has a significantly positive effect on new college graduates' starting salaries at the 1% level. In explicit, graduates who would migrate for work enjoy a 20.3% salary premium over those who did not migrate. For the regression using two instrumental variables, the results show that work migration has a significantly positive effect on new college graduates' starting salaries at the 5% level. In specific, college graduates who would

migrate to work have a 15.8% higher starting salary compared to those who would not migrate to work.

Based on previous research results from the literature review and findings from this study, it is obvious that work migration, student characteristics, family background, academic performance, and institution attributes all have an impact on students' early labor market outcomes. The following results are based on the regression using two instrumental variables. Consistent with previous literature, this study shows that female graduates are at a disadvantage in China's labor market. In more details, female students have an 8.02% lower starting monthly salary than their male counterparts. Being a student leader in high school is associated with a 2.68% increase in starting salary. Similarly, being on the humanities track in high school is associated with a 5.36% decrease in starting salary. A one-unit increase in the student's SES score is associated with a 3.27% increase in starting monthly salary. For a ten-unit increase in NCEE score, we would expect to see about a 5.65% increase in starting salary.

In terms of college experience and institution attributes, having a major in economics or management is associated with an 8.15% decrease in stating monthly salary. A one-unit increase in the preference degree towards one's major is associated with a 2.38% increase in the graduate's starting salary. Having Passed the CET4 and CET6 is associated with a 6.75% and 12.5% increase respectively in starting monthly salary. Being a student leader in college is associated with a 3.35% increase in the new graduate's starting salary. CCP members have a 4.71% higher starting salary than those who are not CCP members. For college work experience, having worked in college is associated with a 4.44% decrease in a graduate's starting salary. Students graduating from a comprehensive institution will have a 7.88% higher salary compared with those graduating from institutions with other concentrations. Graduates studying in a 985

institution will have a 14.8% higher starting salary compared with those in non-key schools. Similarly, studying in a 211 institution is associated with an 11.3% increase in the new graduate's starting salary.

Section 7.2 Significance of This Study

This dissertation is one of the first comprehensive empirical studies to examine the determinants of college graduates' migration and its impact on starting salaries in China. Given the importance of the human capital accumulation in regional development, identifying the determinants and consequences of college graduates' migration is crucial to both education researchers and policy makers. This study creates a conceptual framework and it would substantially sharpen the understanding of the economics of migration from a theoretical perspective.

Second, this study employs rigorous quantitative methodologies to examine the determinants of college graduates' migration decision and its impact on college graduates' labor market outcomes. It is the first one that incorporates student characteristics, institution characteristics and regional characteristics (both economic and non-economic factors) in the regression analysis on determinants of work migration in the Chinese context. In terms of impact on starting salaries, a few empirical studies simply used OLS regressions to estimate it. Their estimates might be biased because work migration is endogenous to college graduates' labor market outcomes. This dissertation study incorporates different identification strategies (IV & PSM) to address the endogeneity problem. This study is more comprehensive than previous studies in terms of identification strategies. In addition, very few Chinese studies on college graduates' labor market outcomes recognized the sample selection issue in the wage equation. This study addressed this problem with the Heckman correction technique.

Third, this dissertation uses the most recent representative sample of the college graduates in China, which makes its conclusions more easily to be inferred to the related population. Furthermore, the findings on China should be of interest to education researchers and policymakers in other countries, because of the rising importance of China in the global community.

Last but not least, the policy significance of this study is evident. It will inform policy makers by deepening the understanding of work migration behavior of college graduates in China. It might help local governments design valid incentives to keep new graduates from flowing out or to recruit new graduates from other areas. Migrants who have high levels of productivity and who adapt rapidly to conditions in the host province's labor market can make a significant contribution to economic growth. The choice of the "right" migration policy can have a significant impact on economic activity both in the short run and in the long run. With more information about the characteristics and backgrounds of migrating college graduates, policy makers would have much insight into the probable effects on the graduates' "mix" within their provincial borders or other important educational considerations. The analysis on college graduates' employment from the perspective of their migration among regions helps understand the conflicts between supply and demand in different regions and provides reference and proof for policy makers to solve the employment problem and improve regional distribution of higher education.

Section 7.3 Limitations and Suggestions for Future Research

This section discusses the limitations of this study and provides suggestions for future empirical research on this topic.

7.3.1 Limitations

This dissertation study has the following limitations.

First, the dataset only provides the salary information for those college graduates who found jobs and reported their salary before their graduation (CSLM 2011 survey was conducted in late May to mid-June). At that time, about one-third of the students in the sample had not been offered a job yet. There is no salary information available for those who found jobs after the survey and therefore these students were not included in the analysis. In addition, using starting salaries as indicator of labor market outcomes has its limitations. There might be some non-monetary benefits associated with the job but were not reflected in the starting salaries. Therefore, the dependent variable of starting salary used in this study may not capture the full picture of these students' early labor market performance. All the above factors might lead to some bias in reporting the conclusions.

Second, the dissertation study was conducted within a cross-sectional framework and it could not capture the long-term impact of migration on college graduates' labor market outcomes. As indicated by Greenwood (1985), longitudinal study with appropriate time-series data on migration and other variables may prove particularly useful in analyzing the determinants and consequence of migration, because they permit a distinctly different approach to the problem of sample selection (i.e., longitudinal data permit researchers to control more directly for unobserved variables that affect earnings and that are correlated with the migration decision). Recognizing this data limitation, the Institute of Education at Tsinghua University is currently working on conducting follow-up surveys on the previously-surveyed college graduates. We hope the dataset will be available in the near future.

Third, the IV coefficients are estimated based on the subsample whose work migration decisions were influenced by their peers' work migration behaviors. These students may be different from other students and they may be more susceptible to institution/provincial migration climate, as they do not have a clear incentive of work migration. Therefore the estimated LATE impacts may be only applicable to a small group of students and all the results can only be interpreted with caution as conditional upon existing conditions.

Fourth, potential validity problems may arise as a result of the missing data and measurement errors in the self-reported data. Though the missing data problem in the CSLM 2011 dataset is not serious overall, some variables had a more than 10% missing rate. In this study, observations with missing dependent variables were deleted from the regression analysis and the "Dummy Flag" strategy was employed to deal with the missing values in covariates. We were not able to employ the multiple imputation strategy because the psmatch2 command in STATA 12 cannot incorporate sample weights. Besides the missing data problem, measurement error is another problem with self-reported data. For instance, students may exaggerate their number of job offers and starting salaries. This may also bias the estimates of the impact of work migration on starting salaries in the analysis.

7.3.2 Suggestions for Future Research

This study provides some preliminary findings on the determinants of college graduates' work migration decisions and its impact on starting salaries in the Chinese context. Based on this research, further studies could contribute to the scholarship in this field in the following three aspects.

First, when analyzing the determinants of college graduates' migration behavior, if more information about possible choice-specific variables (work locations) is available, researchers

should add these variables to the analysis. Researchers should also try to employ the mixed logit model to account for the random taste heterogeneity across individuals. Accounting for this heterogeneity enables estimation of unbiased estimates of individual preferences and enhances the accuracy and reliability of estimates (Greene, 1997). Furthermore, accounting for heterogeneity enables prescription of policies that take equity concerns into account (Gottlieb & Joseph, 2006). An understanding of who will be affected by a policy change in addition to understanding the aggregate economic value associated with such changes is necessary (Boxall and Adamowicz, 2002).

Second, longitudinal study with appropriate time-series data on migration and other variables should be conducted to analyze the determinants and consequence of college graduates' work migration. Students could be surveyed twice during the four years in college and at 6-month, one-year, and three-year out of college. This way, the surveys would well capture information about students' college experience and provide better measurements on college graduates' labor market performance.

Third, the current study employs merely a quantitative approach without qualitative analysis. It would be better to incorporate the qualitative perspective for the future analysis, or to employ a sequential explanatory mixed-method research design that involving both quantitative and qualitative methods. This could help researchers gain an in-depth understanding of the factors that determine college graduates' work migration decision and its impact on starting salaries, and to provide policy makers with more comprehensive evidence to generate effective policies about work migration.

Section 7.4 Policy Implication

Since the beginning of economic reform, the scale of China's higher education has increased continually and it transformed from an elite education to a mass form. In the long term, higher education expansion would raise the entire nation's educational level, which would enable China to compete in the environment of globalization and the knowledge-based economy in the 21st century. However, as the number of college graduates increases, the unemployment issue becomes more severe. Lack of diversity in curricula at different levels and in different divisions of higher education determined that graduates lacked the specialty and the flexibility to respond to market demand. Attitude of graduates to jobs is another critical factor contributing to graduate unemployment. College graduates flocked to big cities and made the unemployment issue worse. This dissertation explores the determinants of college graduates' migration decision and its impact on starting salaries in China. Though the findings need to be examined by future studies, they provide some policy implications.

First, the study reveals the determinants of college graduates' migration decision. The regression analyses in previous sessions show that many individual characteristics are affecting college graduates' migration. Therefore, when any province tries to recruit new college graduates from another province where they graduate, they could target some specific groups and provide incentives for these groups. For example, the study results show that students who are from 985 and 211 institutions are more likely to migrate to work. In such case, recruiting province could put emphasis on these institutions and introduce incentive policies that attract these graduates and simplify "hukou" transfer procedures to help them settle in the new location.

Meanwhile, provincial characteristics (provincial GDP per capita and provincial population) also have significant influence on college graduates' migration., suggesting an important fact that

the development of a province provides a welcome environment to college graduates and sources of their employment. Policies that nurture the provincial economic development and population growth are crucial to attract high skilled people. Furthermore, our results show that, provincial unemployment rate does not have a direct significant impact on graduates' migration, indicating that migrants do not simply focus on the current unemployment rate but broader economic opportunity and development potential are the greater attraction.

Second, cultural factor variable (number of higher education institutions) in our regressions does not have significant influence on college graduates' migration, but ECI scores show significant influence on migration. This result shows that in this relatively early stage of college graduates' migration, though they are aware of the provincial environment quality, the overwhelming concerns of the graduates are career opportunities rather than cultural perspectives of life.

Third, career services center should provide appropriate career guidance based on heterogeneous characteristics of college graduates. For example, students who had study migration behavior before are more likely to migrate again for jobs. Those who did not have study migration would more likely to stay. It would be better to create a student information database and relate such information with career services platform. This would achieve more significant results with limited resources.

Fourth, this dissertation finds that work migration has a significant positive impact on college graduates' early labor market performance-- salaries. This suggests that institutions could use this fact and encourage graduates to conduct job search at a national level. As Niu (2002) indicated, the key to resolve the unemployment is to lower graduates' job expectation. Currently, students from big cities don't want to go to the regions ranked as third or fourth tier; those from

the less developed regions would like to use higher education as a stepping stone for upward mobilization (Wei, 2002). Institutions should instruct students that work migration is a way to accumulate human capital and it brings monetary returns correspondently. Staying in big cities might not be a wise choice for everyone.

Fifth, China has a long history of regional disparities, and disparities in economic development are paralleled by disparities in higher education. Top universities are all located in the economically developed regions in China. This dissertation shows that regional characteristics have significant impact on graduates' migration. In order to resolve the college graduates' unemployment problem fundamentally, it's crucial to promote balanced regional development in both economics and higher education. We believe that the challenges posed by the college graduates' unemployment issue may turn out to bring new opportunities to further reform higher education.

REFERENCES

Abayomi, K., Gelman, A., & Levy, M. (2008). Diagnostics for multivariate imputations. *Applied Stiatistics, 57*(3), 273-291.

Abbott, W. F. S., C. F. (1975). University Prestige and First-Time Undergraduate Migration in the United States. *Sociology of Education, 48*(2), 168-185. doi: 10.2307/2112474

Adelman, C. (2004). Principal Indicators of Student Academic Histories in Postsecondary Education, 1972-2000. Washington, DC: U.S. Department of Education.

Allison, P. D. (2001). Missing Data. Thousand Oaks: Sage.

Alm, J., & Winters, J. V. (2009). Distance and Intrastate College Student Migration. *Economics of Education Review, 28*, 728-738. doi: 10.1016/j.econedurev.2009.06.008

Anderson, W. A. (1928). Social Mobility among College Graduates. *Journal of Educational Sociology, 1*(10), 626-636. doi: 10.2307/2961792

Angrist, J. D., & Pischke, J. (2008). Mostly Harmless Econometrics: An Empiricist's Companion.

Austin, P. C. (2011). An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate Behavioral Research, 46*(3), 399-424. doi: 10.1080/00273171.2011.568786

Bai, L. (2006). Graduate Unemployment: Dilemmas and Challenges in China's Move to Mass Higher Education. *The China Quarterly, 185*, 128-144. doi: 10.1017/S0305741006000087

Barkley, A. P. (1991). Labor Mobility among Agricultural College Graduates: A Human Capital Approach. *Western Journal of Agricultural Economics*, *16*(2), 315-325.

Berker, A. (2009). The Impact of Internal Migration on Educational Outcomes: Evidence from Turkey. *Economics of Education Review, 28*, 739-749. doi: 10.1016/j.econedurev.2009.03.003

Bhat, C. (2001). Quasi-random Maximum Simulated Likelihood Estimation of the Mixed Multinomial Logit Model. *Transportation Research B, 35*, 677-693. doi: 10.1016/S0191-2615(00)00014-X

Bhat, C. (2003). Simulation Estimation of Mixed Discrete Choice Models Using Randomized and Scrambled Halton Sequences. *Transportation Research B, 37*, 837-855. doi: 10.1016/S0191-2615(02)00090-5

Birol, E., Karousakis, K., & Koundouri, P. . (2006). Using a Choice Experiment to Account for Preference Heterogeneity in Wetland Attributes: The Case of Cheimaditida wetland in Greece. Paper presented at the Third World Congress of Environmental and Resource Economists, Kyoto, Japan.

Blundell, R., & Costa-Dias, M. . (2009). Alternative Approaches to Evaluation in Empirical Microeconomics. *Journal of Human Resources, 44*(3), 565-640. doi: 10.1353/jhr.2009.0009

Board, C. (2004). Education Pays 2004: The Benefits of Higher Education for Individuals and Society. Washington, DC: College Board.

Borjas, G. J., Bronars, S. G., and Trejo, S. J. (1992). Self-Selection and Internal Migration in the United States. *Journal of Urban Economics, 32*(2), 159-185. doi: 10.1016/0094-1190(92)90003-4

Borjas, G. J. (1994). The Economics of Immigration. *Journal of Economic Literature, 32*(4), 1667-1717.

Carbone, R. F. (1973). The Future of the Low-Tuition System. *Educational Record, 54*(4), 265-270.

Chen, A., & Coulson, N. E. (2002). Determinants of urban migration: Evidence from Chinese cities. *Urban Studies, 39*(12), 2189-2197. doi: 10.1080/0042098022000033818

Chen, B. Y., Y. (2010). Developing students' leadership: Based on career choice courses. *Management of Elementary and Secondary Schools, 7*, 17-18.

Chinese Department of Education (2008). China Education Statistics. China Education Statistics Retrieved 2012/7/2, http://www.stats.edu.cn/sjcx.aspx#

Christal, M. E. (1982). The Sweep the South: Fact or Fallacy. Paper presented at the Annual Conference of the Southern Association for Institutional Research, Birmingham, Alabama.

Ciuk, D. J., & Pyle, K. (2009). Doing more with missing data: Incorporating systematic missingness into survey research. Paper presented at the Annual Meetings of the Midwestern Political Science Association. Chicago, IL.

Clark, D. E., & Hunter, W. J. (1992). The Impact of Economic Opportunity, Amenities and Fiscal Factors on Age-Specific Migration Rates. *Journal of Regional Science, 32*(3), 349-365. doi: 10.1111/j.1467-9787.1992.tb00191.x

Cobb-Clark, D. A. (1990). Immigrant Selectivity: The Roles of Household Structure and U.S. Immigration Policy. (Ph.D.), University of Michigan.

DaVanzo, J. (1978). Does Unemployment Affect Migration? Evidence from Micro Data. *Review of Economics and Statistics, 6*(4), 504-514. doi: 10.2307/1924242

Detang-Dessendre, C., & Molho, I. (2000). Residence Spells and Migration: A Comparison for Men and Women. *Urban Studies, 37*(2), 247-260. doi: 10.1080/0042098002177

Du, G., & Yue, C. . (2010). Influencing factors research of the university graduates employment opportunity (in Chinese). *China Higher Education Research, 10*, 67-70.

Duan, C., Yang, K., Zhang, F., & Lu, X. . (2008). Nine Trends of Changes in Migration Population Since Chinese Economic Reform and Opening Up. *Population Research, 32*(6), 30-43.

Enders, C. (2010). Applied Missing Data Analysis. New York: Guilford Press.

Faggian, A., McCann, P., & Sheppard, S. (2007). Some Evidence That Women Are More Mobile Than Men: Gender Differences in U.K. Graduate Migration Behavior. *Journal of Regional Science, 47*(3), 517-539. doi: 10.1111/j.1467-9787.2007.00518.x

Fenske. R. H., Scott., C. S., & Carmody, J. F. (1974). Recent Trends in Studies of Student Migration. *The Journal of Higher Education, 45*(1), 61-74. doi: 10.2307/1980649

Ferriss, A. L. (1973). Higher Educational Institutions and the Migration of Talent. *College and University, 49*(1), 19-29.

Fielding, A., & Halford, S. (1993). Geographies of Opportunity: A Regional Analysis of Gender-Specific Social and Spatial Mobilities in England and Wales, 1971-81. *Environment and Planning, 25*, 1421-1440. doi: 10.1068/a251421

Florida, R. (2002). The Rise of the Creative Class. New York, NY: Basic Books.

Frees, E. W. (1992). Forecasting State-to-State Migration Rates. *Journal of Business and Economic Statistics, 10*(2), 153-167.

Gabriel, S. A., Shack-Marquez, J., & Wascher, W. L. (1993). The Effects of Regional House Price and Labor Market Variability on Interregional Migration: Evidence from the 1980s. In T. K. a. M. A. Turner (Ed.), Housing Markets and Residential Mobility. Washington: Urban Institute Press.

Glaeser, E. L., & Albert, S. (2003). The Rise of the Skilled City. Cambridge, MA: National Bureau of Economic Research.

Gossman, C. S. (1968). Migration of College and University Students in the United States. Seattle, University of Washington Press.

Gottfried, A. E., Gottfried, A. W., Reichard, R. J., Guerin, D. W., Oliver, P. H., & Riggio, R. E. . (2011). Motivational roots of leadership: A longitudinal study from childhood through adulthood. *The Leadership Quarterly, 22*(3), 510-519. doi: 10.1016/j.leaqua.2011.04.008

Gottlieb, P. D., & Fogarty, M. (2003). Educational Attainment and Metropolitan Growth. *Economic Development Quarterly, 17*, 325-336. doi: 10.1177/0891242403257274

Gottlieb, P. D., & Joseph, G. (2006). College-to-Work Migration of Technology Graduates and Holders of Doctorates within the United States. *Journal of Regional Science, 46*(4), 627-659. doi: 10.1111/j.1467-9787.2006.00471.x

Greene, W. H. (1997). Econometric Analysis. New Jersey: Prentice Hall.

Greenwood, M. J. (1973). The Geographic Mobility of College Graduates. *The Journal of Human Resources, 8*(4), 506-515. doi: 10.2307/144860

Greenwood, M. J. (1975). Research on Internal Migration in the United States: A Survey. *Journal of Economic Literature, 13*(2), 397-433.

Greenwood, M. J. (1985). Human Migration: Theory, Models, and Empirical Studies. *Journal of Regional Science,* 521-544. doi: 10.1111/j.1467-9787.1985.tb00321.x

Greenwood, M. J., & Hunt, G. L. (1989). Jobs versus Amenities on the Analysis of Metropolitan Migration. *Journal of Urban Economics, 25*(1), 1-16. doi: 10.1016/0094-1190(89)90040-5

Greenwood, M. J., et al. (1991). Migration, Regional Equilibrium, and the Estimation of Compensating Differentials. *American Economic Review, 81*(5), 1382-1390.

Group, U. S. C.). Introduction to Stata. Retrieved June, 14, 2014, from http://www.ats.ucla.edu/stat/stata/dae/mlogit.htm

Guo, C., Tsang, M., & Ding, X. (2010). Gender Disparities in Science & Engineering in Chinese Universities. *Economics of Education Review, 29*(2), 225-235. doi: 10.1016/j.econedurev.2009.06.005

Guo, S., & Fraser, M. W. (2009). Propensity Score Analysis: Statistical Methods and Applications (Vol. 12). Thousand Oaks, CA: Sage Publications, Inc.

Hansen, S., Ban, C., & Huggins, L. (2003). Explaining the "Brain Drain" from Older Industrial Cities: The Pittsburgh Region. *Economic Development Quarterly, 17*, 132-147. doi: 10.1177/0891242403017002002

Heaton, C., & Throsby, D. . (1998). Benefit-Cost Analysis of Foreign Student Flows from Developing Countries: The Case of Postgraduate Education. *Economics of Education Review, 17*(2), 117-126. doi: 10.1016/S0272-7757(97)00019-8

Herzog, H. W., & Schlottmann, A. M. (1984). Labor Force Mobility in the United States: Migration, Unemployment, and Remigration. *International Regional Science Review, 9*, 43-58. doi: 10.1177/016001768400900102

Herzog, H. W., & Schlottmann, A. M. (1986). High-Technology Jobs and Worker Mobility. *Journal of Regional Science 26,* 445-459. doi: 10.1111/j.1467-9787.1986.tb01053.x

Herzog, H. W., & Schlottmann, A. M. (1991). Metropolitan Dimension of High-Technology Location in the U.S.: Worker Mobility and Residence Choice. In H. W. Herzog, & Schlottmann, A. M. (Ed.), Industry Location and Public Policy (pp. 169-189). Knoxville, TN: University of Tennessee Press.

Hicks, J. R. (1932). The Theory of Wages. London: Macmillan.

Hirono, K., & Imbens, G. W. (2001). Estimation of causal effects using propensity score weighting: An application to data on right heart catheterization. *Health Services & Outcomes Research Methodology, 2*, 259-278.

Houweling, T. A., Kunst, A.E., & Mackenbach, J. P. (2003). Measuring health inequality among children in developing countries: Does the choice of the indicator of economic status matter? *International Journal for Equity in Health, 2*(8), 1-12. doi: 10.1186/1475-9276-2-8

Huang, J. (2007). Empirical analysis of the personal characteristics and college students' labor market outcomes. (in Chinese). *Higher Education Exploration, 4*, 117-120.

Imbens, G. W., & Wooldridge, J. M. (2007). Instrumental variables with treatment effect heterogeneity: Local average treatment effects. NBER report: What's New in Econometrics.

Illinois Board of Higher Education (2005). Illinois Higher Education Annual Report: 2005. Springfield: IL.

Institute for Higher Education Policy (2005). The Investment Payoff: A 50-state Analysis of the Public and Private Benefits of Higher Education. Washington, DC: Institute for Higher Education Policy.

Jacobs, A. (2010, December 11, 2010). China's Army of Graduates Struggles for Jobs, *The New York Times*. Retrieved from http://www.nytimes.com/2010/12/12/world/asia/12beijing.html?pagewanted=all&_r=0

Khoury, R. M. (1977). The Geographic Mobility of Academic Talent: Some Evidence from Sociology. *Research in Higher Education 7*(2), 155-165. doi: 10.1007/BF00981753

Kodrzycki, Y. K. (2001). Migration of Recent College Graduates: Evidence from the National Longitudinal Survey of Youth. *New England Economic Review*, January/February.

Kyung, W. (1992). Determinants of In-Migration of College Students to the State of New York. Discussion Paper 92-1. Discussion Paper. Center for Urban Studies, School of Urban and Public Affairs, Portland State University.

Lai, D., Meng, D., & Su, L. (2012). Substitution or complementation: a study on joint mechanism of human capital and social capital to college graduates' employment (in Chinese). *Peking University Education Review, 10*(1), 13-31.

Lankford, F. G., & Taylor, A. L. (1971, Jan. 15, 1971). College and University Migration: The Case of Virginia, *University of Virginia Newsletter*, pp. 17-20.

Lee, E. S. (1966). A Theory of Migration. *Demography, 3*(1), 47-57. doi: 10.2307/2060063

Li, F., Liu, F., & Guo, Z. (2009). An Empirical Study of Migration in Postgraduate Employment. *Tsinghua Journal of Education, 30*(4), 67-71.

Li, F., Zhao, Y., & Guo, Z. (2010). Empirical Research on the Income of Postgraduate's Employment Migration. *Research in Higher Education of Engineering, 3*, 60-65.

Li, F., & He, G. (2011). Pull and Push: The Dual-drive of Migration Employment for Post-graduates. *Journal of Higher Education, 32*(4), 25-29.

Li, H., Meng, L., Shi, X., & Wu, B. . (2012). Does attending elite colleges pay in China? *Journal of Comparative Economics, 40*(1), 78-88. doi: 10.1016/j.jce.2011.10.001

Li, J., & Lang, W. (2012). The significance and strategies to foster leadership of high school studnets. *Research in Teaching, 35*(4), 111-113.

Li, L., & Tian, G. (2013). Girls "incoming", Female Flourishes while Male Declines in College? Retrieved 2014-5-8, 2014, from http://gaojiao.jyb.cn/gjsd/201310/t20131017_555945.html

Li, S., & Xing, C. (2010) China's Higher Education Expansion and its Labor Market Consequences.

Li, W., et al. (2002). The Situation of China's Graduate Employment and Our Suggestions. *University Administration, 8*, 21.

Little, R. J., & Rubin, D. (2002). Statistical Analysis with Missing Data (2nd Edition ed.). New York: John Wiley & Sons, Inc.

Loewen, P. J. (2010). Constant Causal Effects Assumption. In A. J. Mills, Durepos, G., & Wiebe, E. (Ed.), *Encyclopedia of Case Study Research*, SAGE Publications, Inc.

Long, J. E. (1980). The Effect of Americanization on Earnings: Some Evidence for Women. *Journal of Political Economics, 88*(3), 620-629. doi: 10.1086/260892

Lu, D. (2008). The emergency and development of adolescent leadership--An analysis based on interviews with 26 high school students in 3 cities. *China Youth Study, 5*, 10-15.

Mathur, V. (1999). Human Capital Based Strategy for Regional Economic Development. *Economic Development Quarterly, 13*, 203-216. doi: 10.1177/089124249901300301

McConnell, C. R., Brue, S. R., & Macpherson, D. A. (2010). Contemporary Labor Economics. New York: McGraw-Hill.

McFadden, D., & Train, K. (2000). Mixed MNL Models of Discrete Response. *Journal of Applied Econometrics, 15*, 447-470. doi: 10.1002/1099-1255(200009/10)15:5<447::AID-JAE570>3.0.CO;2-1

McHugh, M. L. (2009). The odds ratio: Calculation, usage, and interpretation. *Biochemia Medica, 19*(2), 120-126.

McHugh, R., & Morgan, J. N. (1984). The Determinants of Interstate Student Migration: a Place-to-place Analysis. *Economics of Education Review, 3*(4), 269-278. doi: 10.1016/0272-7757(84)90045-1

McPherson, M. S., & Schapiro, M. O. (1991). Does Student Aid Affect College Enrollment? New Evidence on a Persistent Controversy. *The American Economic Review, 81*(1), 309-318.

Miao, M., & Ding, Y. (2009). China Moves to Solve Graduate Unemployment Issue. Retrieved 2012/7/12, from http://news.xinhuanet.com/english/2009-01/11/content_10639139.htm

Mixon, F. G. (1992). Factors Affecting College Student Migration across States. *International Journal of Manpower, 13*(3), 63-68.

Mixon, F. G., & Hsing, Y. (1994). College Student Migration and Human Capital Theory: A Research Note. *Education Economics, 2*(1), 65-74. doi: 10.1016/S0272-7757(05)80056-1

Mixon, F. G., & Hsing, Y. (1994). The Determinants of Out-of-State Enrollments in Higher Education: A Tobit Analysis. *Economics of Education Review, 13*(4), 329-335. doi: 10.1080/09645299400000005

Murray, M. P. (2006). Avoiding invalid instruments and coping with weak instruments. *Journal of Economic Perspectives, 20*(4), 111-132. doi: 10.1257/jep.20.4.111

MyCOS Institute (2011). The 2011 China Graduate Employment Report

Nakosteen, R. A., & Zimmer, M. (1980). Migration and Income: The Question of Self-Selection. *Southern Economic Journal, 46*, 840-851. doi: 10.2307/1057152

National Center for Public Policy & Higher Education (2005). Policy Alert: Income of US Workforce Projected to Decline if Education doesn't Improve. San Jose: CA: National Center for Public Policy & Higher Education.

Niu, J. (2002). Where the difficulties lie in graduates' unemployment. *References for Policy Makers, 35*, 12-13.

Normand, S. L. T., Landrum, M.B., Guadagnoli, E., Ayanian, J.Z., Ryan, T.J., Cleary, P.D., & McNeil, B.J. (2001). Validating recommendations for coronary angiography following an acute myocardial infarction in the elderly: A matched analysis using propensity scores. *Journal of Clinical Epidemiology, 54*, 387-398.

Pearson, K. (1901). On lines and planes of closest fit to systems of points in space. *Philosophical Magazines, 2*(11), 559-572. doi: 10.1080/14786440109462720

Perry, K. K. (2001). Where College Students Live After They Graduate.

Pissarides, C. A., & McMaster, I. (1990). Regional Migration, Wages and Unemployment: Empirical Evidence and Implications for Policy. *Oxford Economic Papers, 42*, 812-831.

Qian, X. (2000). Is the idea of using education to stimulate economy too subjective? *On China's Educational Policy*, 190-191.

Qing, S., & Zeng, X. (2009). Employability, internship and graduate employment: Based on Shandong survey data. (In Chinese). *Chinese Journal of Population Science, 6*, 102-108.

Ravenstein, E. J. (1885). The Laws of Migration. *Journal of the Royal Statistical Society, XLVIII*(2), 167-227.

Ren, N., Guo, J., & Pan, F. (2013). College students' part-time working experience and the competitiveness in the labor market--using Beijing Normal University as an example. (in Chinese). *Money China, 15*, 266-268.

Reseck, et. al. (2000). Illinois Higher Education: Building the Economy, Shaping Society. Champaign-Urbana, IL: Institute of Government and Public Affairs, University of Illinois.

Rozelle, S., Taylor, J. E., & deBrauw, A. . (1999). Migration, Remittances, and Agricultural Productivity in China. *The American Economic Review, 89*(2), Papers and Proceedings of the One Hundred Eleventh Annual Meeting of the American Economic Association (May, 1999)), 287-291.

Rubin, D. B. (1976). Inference and missing data. *Biometrika, 63*, 581-592.

Rubin, D. B. (1987). Multiple Imputation for Nonresponse in Surveys. New York: John Wiley & Sons.

Schlottmann, A. M., & Herzog, H. W. (1981). Employment Status and the Decision to Migrate. *Review of Economics and Statistics, 63*(4), 590-598. doi: 10.2307/1935855

Schneider, B., & Paul, M. (1999). Understanding high school student leaders: Predicting teacher ratings of leader behavior. *The Leadership Quarterly 10*(4), 609-636.

Schultz, T. W. (1962). Reflections on Investments in Man. *Journal of Political Economy, 70*(5), 1-8. doi: 10.1086/258723

Schultz, T. W. (1971). Investment in Human Capital: The Role of Education and of Research. New York: Free Press.

Scott, D. M., Coomes, P. A., & Izyumov, A. I. (2005). The Location Choice of Employment-based Immigrants Among U.S. Metro Areas. *Journal of Regional Science, 45*(1), 113-145. doi: 10.1111/j.0022-4146.2005.00366.x

Shi, X., Li, H., Meng, L., & Wu, B. (2012). Analysis on the starting salary difference between elite and non-elite graduates Study, development, employment: Research papers on Chinese College Student Survey (pp. 278-300). Beijing: Tsinghua University.

Sjaastad, L. A. (1962). The Costs and Returns of Human Migration. *Journal of Political Economy, 70*(5), 80-93. doi: 10.1086/258726

Smith, R. L., & Wall, A. F. (2006). Estimating the Economic Impact of College Student Migration from Illinois. Carbondale, IL: Illinois Education Research Council, Southern Illinois University.

Staiger, D., & Stock, J. H. . (1997). Instrumental variables regression with weak instruments. *Econometrica, 65*(3), 557-586.

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

Tassinopoulos, A., & Werner, H. . (1999). To Move or Not to Move--Migration of Labour in the European Union. *IAB Labour Market Research Topics, 35*, 1-19.

Titus, M. A. (2010). Understanding the relationship between working while in college and future salaries. In L. W. Pern (Ed.), Understanding the working college students: New research and its implications for policy and practice (Vol. 1st ed, pp. 261-283). Sterling, VA.: Stylus Publishing, LLC.

Todaro, M. P. (1969). A Model of Labor Migration and Urban Unemployment in Less Developed Countries. . *The American Economic Review, 59*(1), 138-148.

Tornatzky, L. G., et. al. (2001). Who Will Stay and Who Will Leave? Individual, Institutional and State-Level Predictors of State Retention of Recent Science and Engineering Graduates. Raleigh-Durham, NC: Southern Growth Policies Board, Southern Technology Council.

Train, K. (2000). Halton Sequences for Mixed Logit Working Paper No. E00-278: Department of Economics, University of California, Berkeley.

Train, K. (2003). Discrete Choice Methods with Simulation. New York: Cambridge University Press.

Treyz, G. I., Rickman, D. S., Hunt, G. L., & Greenwood, M. J. (1993). The Dynamics of U.S. Internal Migration. *Review of Economics and Statistics, 75*(2), 209-224. doi: 10.2307/2109425

Trow, M. (2005). Reflections on the Transition from Elite to Mass to Universal Access: Forms and Phases of Higher Education in Modern Societies since WWII. In P. Altbach (Ed.), International Handbook of Higher Education (pp. 1-66): Kluwer.

Vyas, S., & Kumaranayake, L. . (2006). Constructing socio-economics status indices: How to use principal components analysis. *Health Policy and Planning, 21*(6), 459-468. doi: 10.1093/heapol/czl029

Wei, L., et al. (2002). The situation of China's graduate employment and our suggestions. *University Administration, 8*, 21-26.

Western Interstate Commission of Higher Education (2005). Student Migration: Relief Valve for State Enrollment and Demographic Pressures. Boulder: CO.

Wooldridge, J. M. (2001). Applications of generalized method moments estimation. *Journal of Economic Perspectives, 15*(4), 87-100. doi: 10.1257/jep.15.4.87

Xie, Y., & Li, X. (2010). The research on status and determinants of job search duration of college graduates: an empirical research based on survival model. *Peking University Education Review, 8*(2), 158-168.

Yan, F., & Mao, D. (2008). The impact of social capital on employment of university graduates. *Fudan Education Forum, 6*(4), 56-65.

Yang, K. (2014). 2014 Report on China's Ecology Civilization Index. *China Economics Weekly*. Retrieved 7/20, 2014, from http://www.ceweekly.cn/2014/0716/87178.shtml

Yang, P. (2011). The impact of financial aid on learning, career decisions, and employment. *Chinese Education & Society, 44*(1), 27-57. doi: 10.2753/CED1061-1932440102

Yousefi, M., & Rives, J. (1987). Migration Behavior of College Graduates: An Empirical Analysis. *Journal of Behavioral Economics, 16*, 35-49. doi: 10.1016/0090-5720(87)90037-4

Yue, C., & Zhou, J. . (2005). Why Do the College Graduates Choose Inter-province Employment? *Tsinghua Journal of Education, 26*(2), 34-41.

Yue, D. (2010). Factors affecting employment opportunities for college graduates. *China Higher Education Research, 11*, 68-71.

Zeng, D. Z., & Wang, S. (2007). China and the Knowledge Economy: Challenges and Opportunities., World Bank Policy Research Working Paper 4223.

Zhang, G., Zhao, Y., & Lei, J. (2012). Expansion of higher education and the employment crisis: policy innovations in China. *On the Horizon, 20*(4), 336-344.

Zhao, Y. (1997). Labor Migration and Returns to Rural Education in China. *American Journal of Agricultural Economics, 79*(4), 1278-1287. doi: 10.2307/1244284

Zhao, Y. (1999). Leaving the Countryside: Rural-to-Urban Migration Decisions in China. *The American Economic Review, 89*(2, Papers and Proceedings of the One Hundred Eleventh Annual Meeting of the American Economic Association (May, 1999)), 281-286.

Zhu, N. (2001). The China's Population Transfer under the Four-dimension Economy. *Population & Economics, 1*, 44-52.

APPENDIX

Appendix 1. SPSS outputs of the construction of the Index of Socio-economic Status

Table A1.1 Descriptive Statistics

	Mean	Std. Deviation	Number of Observations
Log(household income) /lnfaminc	10.52	0.835	5,231
Log(residency area) /lnresarea	4.67	0.397	5,231
Mother's years of schooling /momyrsch	9.75	3.892	5,231
Father's years of schooling /dadyrsch	10.90	3.395	5,231
Residency at rural /resrural	0.46	0.499	5,231
Residency in ordinary commercial building /resordinary	0.25	0.434	5,231
One parent is manager /hous_manager	0.14	0.351	5,231
One parent is professional /hous_professional	0.17	0.374	5,231
One parent is ordinal staff /hous_ordstaff	0.16	0.364	5,231
One parent is farmer or worker /hous_farmworker	0.48	0.500	5,231
One parent works in government /hous_gov	0.10	0.297	5,231
One parent works in public institutes /hous_inst	0.19	0.389	5,231
One parent works in public service industry (edu. & medicine) /hous_pub	0.15	0.360	5,231
One parent works in service or retail industry /hous_sersale	0.25	0.431	5,231

Table A1.2 KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy		.805
Bartlett's Test of Sphericity	Approx. Chi-Square	20618.685
	Df	91
	Sig.	.000

Table A1.3 Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	4.249	30.351	30.351	4.249	30.351	30.351
2	1.647	11.763	42.114	1.647	11.763	42.114
3	1.326	9.475	51.588	1.326	9.475	51.588
4	1.033	7.380	58.968	1.033	7.380	58.968
5	1.004	7.169	66.137	1.004	7.169	66.137
6	.780	5.574	71.711			
7	.716	5.116	76.827			
8	.643	4.593	81.420			
9	.603	4.307	85.727			
10	.558	3.984	89.711			
11	.382	2.725	92.436			
12	.375	2.679	95.115			
13	.371	2.652	97.767			
14	.313	2.233	100.000			

Extraction Method: Principal Component Analysis

Table A1.4 Component Matrix

	Component				
	1	**2**	**3**	**4**	**5**
lnfaminc	.625	-.159	.067	.067	.415
lnresarea	-.120	.244	.377	.597	.562
Mother's years of schooling	.723	-.019	.064	-.170	.049
Father's years of schooling	.729	.065	.139	-.099	.004
resrural	-.739	.333	.107	.202	-.053
resordinary	.514	-.397	-.109	-.319	.317
hous_manager	.568	-.001	.577	-.040	-.110
hous_professional	.541	.517	-.324	-.026	.088
hous_ordstaff	.307	-.396	-.210	.464	-.466
hous_farmworker	-.602	.153	.010	-.275	-.001
hous_gov	.414	-.059	.633	.093	-.365
hous_inst	.606	.439	-.261	.200	-.131
hous_pub	.582	.571	-.234	.077	-.109
hous_servsale	.168	-.565	-.357	.378	.105

Extraction Method: Principal Component Analysis; 5 components extracted.

Appendix 2. OLS Estimates of the Impact of Work Migration on Starting Salaries by Gender

Table A2.1 OLS Estimates of the Impact of Work Migration on Starting Salaries by Gender (Dependent variable: monthly starting salary in log form)

	(1) Female	(2) Male
Work Migration	0.062*	0.128***
	(0.040)	(0.036)
Age	0.023	-0.006
	(0.016)	(0.011)
Minority	0.147*	-0.078
	(0.074)	(0.042)
Single child	0.007	0.001
	(0.036)	(0.029)
From rural area	-0.102***	-0.001
	(0.038)	(0.032)
SES score	0.026	0.012
	(0.022)	(0.017)
Student leader in high school	0.062*	0.064**
	(0.027)	(0.022)
Humanity track in high school	-0.078	-0.057
	(0.043)	(0.051)
Arts or sports track in high school	0.067	-0.086
	(0.075)	(0.096)
NCEE score	0.010***	0.004
	(0.003)	(0.002)
Average course score	0.003	-0.002
	(0.003)	(0.002)
Science or engineering major	-0.013	-0.118*
	(0.044)	(0.051)
Economics or management major	-0.118**	-0.184***
	(0.044)	(0.050)
Have a minor	0.011	-0.044
	(0.047)	(0.041)
Preference degree of one's major	0.025	0.007
	(0.018)	(0.013)
Pass CET-4	0.103*	0.043
	(0.047)	(0.029)
Pass CET-6	0.160***	0.165***
	(0.048)	(0.038)
Student leader in college	0.035	0.033

	(0.032)	(0.028)
CCP member	0.034	0.022
	(0.033)	(0.025)
Have professional certificates	0.001	-0.033
	(0.029)	(0.022)
Have worked in college	0.039	-0.059*
	(0.047)	(0.029)
Have merit-based aid	0.044	0.044
	(0.031)	(0.028)
Have need-based aid	0.0001	-0.013
	(0.032)	(0.028)
Have loan	-0.029	-0.015
	(0.030)	(0.025)
Comprehensive institutions	0.021	0.170**
	(0.051)	(0.058)
Engineering-concentrated institutions	0.059	0.060
	(0.039)	(0.046)
985' institution	0.156**	0.079
	(0.060)	(0.052)
211' insititution	0.107***	0.127***
	(0.033)	(0.028)
Independent college	-0.362	-0.058
	(0.193)	(0.049)
Institution located in central or west area	0.073	0.011
	(0.040)	(0.026)