FAMILY SOCIOECONOMIC STATUS, COLLEGE EDUCATION AND EARNINGS IN CHINA

Linri Wang

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 Science

COLUMBIA UNIVERSITY

2005

ABSTRACT

Family Socioeconomic Status, College Education and Earnings in China

Linri Wang

This study examines the effect of family socioeconomic status on college education and college graduates' initial earnings in China. The data set used is the Survey of the Willingness on Higher Education Institution Graduates' Occupational Choice and Employment, 2003 conducted by Peking University, China. Without correcting for self-selection bias due to family socioeconomic status (SES), gender and ethnicity were found to be non-significant in previous studies. Using the Heckman two-stage method to control for self-selection bias, this study finds significant impact of family SES on college graduates' initial earnings through higher education. A significant earnings gap between genders emerges when self-selection bias is corrected, and the gap is as much as 9.3% of the mean annual earnings. However, ethnicity is still non-significant even after the correction of self-selection bias. Institutional characteristics' effects are largely picked up by family SES and become insignificant after correction for self-selection bias.

Using the most current data, this study, for the first time, documents the current situation of family socioeconomics status, especially the latent SES

factors' impact on Chinese college graduates' initial earnings in a quantitative manner and provides a base line for future research. A number of direction for future research is recommended.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGMENTS

I would like to express my gratitude to several individuals, without whom I would not have completed this study. Though this study is accomplished with a very tight schedule and limited resource, there is never a lack of support of my advisors, colleagues and friends. I am thankful to Professor Dongmao Wen and Professor Xiaohao Ding, who generously allowed me to use their latest survey data. Their crucial help makes this study possible. During the study, Professor Dongmao Wen also gave me numerous valuable suggestions and comments. Dr. Helga Marchelli, Dr. Yanqing Ding and Dr. Zeyu Xu all contributed their precious idea and effort to help me complete this study.

I am deeply grateful to my dissertation committee. Every single professor, including Professor Francisco Rivara-Batiz, Professor Kevin Dougherty, Professor Lyle Yorks and Professor Andrew Nathan, spent a lot of time and effort on reading my work and helping me improve the study.

Finally, with all of my respect and thankfulness, I want to thank Professor Mun Tsang, my fatherlier advisor and my forever mentor, who gives

me immense encouragement, leads me through the gloomiest moment of my study and shows me a perfect example of what a good man is. Without him, I could not imagine the achievement I have accomplished.

For Zhongde Lin, Amy Wang, and Yifen Xie
My guiding lights

CHAPTER I

INTRODUCTION

I.1 Statement of the Problem

Education is an important determinant of future earnings and higher earnings lead to better living conditions, health, and well-being (Solomon & Fagnano, 1995; McMahon, 1987). Extensive studies on estimating the economic benefits of schooling in the United States from the 1950s to the 1980s have all come to the same conclusion; everything else being equal, those with more education earn more (Becker, 1993; Taubman and Wales, 1974; Solomon, 1981; Rumberger, 1987; Murphy and Welch, 1989). Most jobs have education requirements for entry and advancement and consequently, schooling attainment became an important determinant of social and occupational mobility (Levin, 1995). In industrialized countries, obtaining higher education has become a gatekeeper to better jobs and middle-class status. Inevitably, higher education becomes an important determinant of earnings and occupational mobility.

Studies of developing and developed countries have suggested that the dispersion of schooling is associated with income inequality (Velloso, 1995; World Bank, 2000). Empirical results in the United States have revealed that schooling level attained is a significant explainer of earnings differences among groups (Carnoy, 1995). As the perceived benefit of a college education has increased, the number of aspirants desiring higher education has also increased. Higher education around the world is growing at a tremendous rate. Enrollments in higher education worldwide doubled in a span of just 20 years from 40.3 million in 1975 to 80.5 million in 1995 (World Bank, 2000). Yet, with the increasing enrollment, there exists substantial inequality in higher education by socioeconomic status (SES) among many societies (World Bank, 2000). For instance, in Latin America, the professional stratum accounts for no more than 15 percent of the general population but their children account for nearly half of the total enrollment in higher education on the continent (World Bank, 2000). This is also true in the United States where minorities are underrepresented at institutions in higher education. There is substantially more African-American, Hispanic, and low socioeconomic students attending public two-year institutions than private or public four-year institutions (NCES, 2004). Extensive literature in the United States have been published on estimating these inequalities and to posit explanations for them mainly because of the detailed census statistics collected

due to the social pressure exerted by the civil rights movement in the 1950s and 60s (Levin, 1995). Carnoy (1995) has indicated that identifying these inequalities is largely a function of data collection politics.

Though higher education is an important determinant of the future earnings of labor, correction for "selection bias" in estimating the education-earnings relationship is sometimes omitted by researchers. In fact, the differences of aptitude among students do affect their college choice and career track, which further influence the job expectation and future earnings by affecting college choice. Since Willis and Rosen's (1979) famous education and self-selection study revealed the positive correlation between "ability bias" and college choice, a number of related studies have found the positive correlation between aptitude and higher education, which means occupational choice and job expectation might not only be affected by students' SES background but also by the aptitude of the students. Considering decision to work or to continue post-secondary education is nonrandom, consequently the working and schooling samples potentially have different characteristics. Sample selection bias arises when some component of the career track decision is relevant to the earnings determining process. That is, when some of the determinants of the schooling / working decision are also influencing earnings. When the relationship between the career decision and the wage is purely through the observables, however, one can control for this by including the

appropriate variables in the wage equation. Thus, sample selection bias will not arise purely because of differences in observable characteristics (Vella, 1998). Therefore, correcting self-selectivity bias for aptitude differences, to what degree the family SES background affects in turn students' college choice and the future earnings becomes an arguable topic. A lot of researchers and social scientists claim that SES yields unequal education opportunity and further influence the life time earnings of a person.

In China, the national economy and higher education institution have undergone a serial of reforms after the Open Policy in late 1970s. The reforms change the relationships between higher education and the labor market, especially the centrally planned job placement system (the *allot* system). One of the major achievements of the higher education reform is the expansion of college enrollment, which largely mitigated the shortage of high skilled labor supply and partially contributed to the abolishment of the obsolete centrally planned job placement system. Since late 1980s, a market oriented labor market and job placement system has been established gradually. The higher education and its relationship with earnings become more and more important to the people's everyday life and national economy.

Good data and studies on higher education and labor market are sorely lacking in China. The deficient data on socioeconomic background can be partly

attributable to the perception that such information is an invasion of privacy and thus is frowned upon. Additionally, the government also assumes that, with the dominant public higher education system, an equitable higher education is achieved. Accordingly, the published data do not reveal the detailed students' family socioeconomic status. However, it is absolutely necessary and crucial that such analysis to be conducted at this time because both China's post-secondary education and labor market are undergoing tremendous changes and is necessary to document these changes now.

Since China is growing fast in national and economic development, inevitably, the socioeconomic structure of China' society could be changing. Some observers issue the warning that the polarizing economic development will result in an acute and serious consequence of inequality among social members. Through higher education, family SES could exacerbate the differentiation by affecting life time earnings (Yu & Lu, 2001). But there is a lack of evidence on this subject. This investigation uses the latest survey data to study the relationship between family socioeconomic status, higher education and earnings.

I.2 Development of Higher Education in China

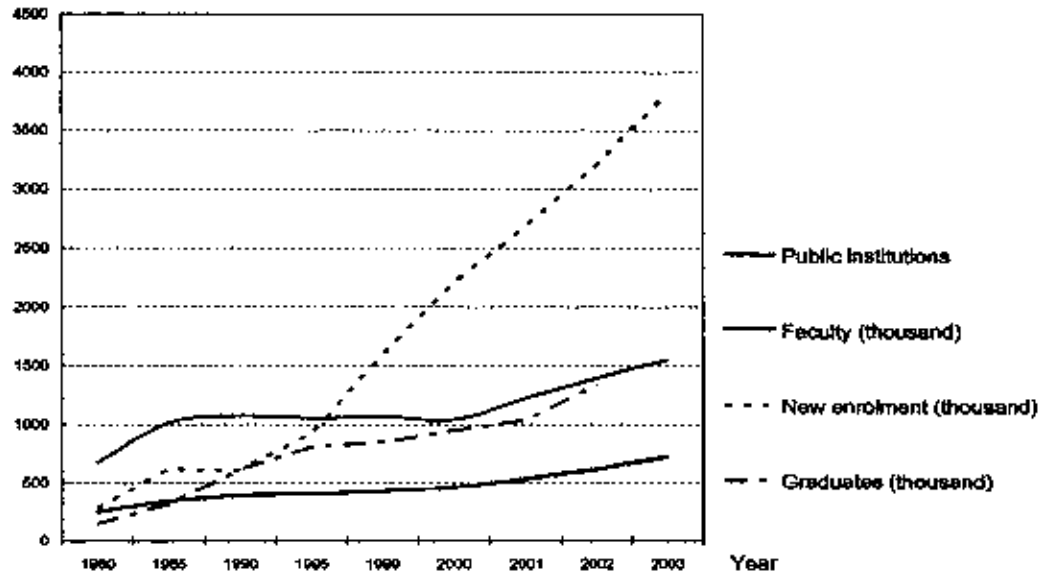
I.2.1 Basic Statistics and Brief Historical Overview

Since the founding of the People's Republic of China in 1949, China's higher education has accomplished a lot (Tsang, 2000). By the year of 2003, the total number of public higher education institution has reached 2110. Among these public 2110 tertiary institutions, 1552 of them are at or above the undergraduate level, and 720 of them can confer graduate degrees (MOE, 2004).

In terms of enrollment, by 2003, total tertiary enrollment topped 19 million for public institutions, including 0.2689 million new graduate level enrollment and 3.82 million new undergraduate enrollment; gross enrollment rate reached 17% of the age cohort. For non-government sector, by 2003, the total number of accredited *Minban* (non-public) post-secondary institution increased to 173, which enrollment 0.81 million students. There are 1104 other *Minban* post-secondary institutions, which are yet to be accredited by the Ministry of Education, enroll 1 million students (MOE, 2004).

China's higher education has made substantial progress, especially in the last two decades. The detailed trends of changes are shown in the following figures.

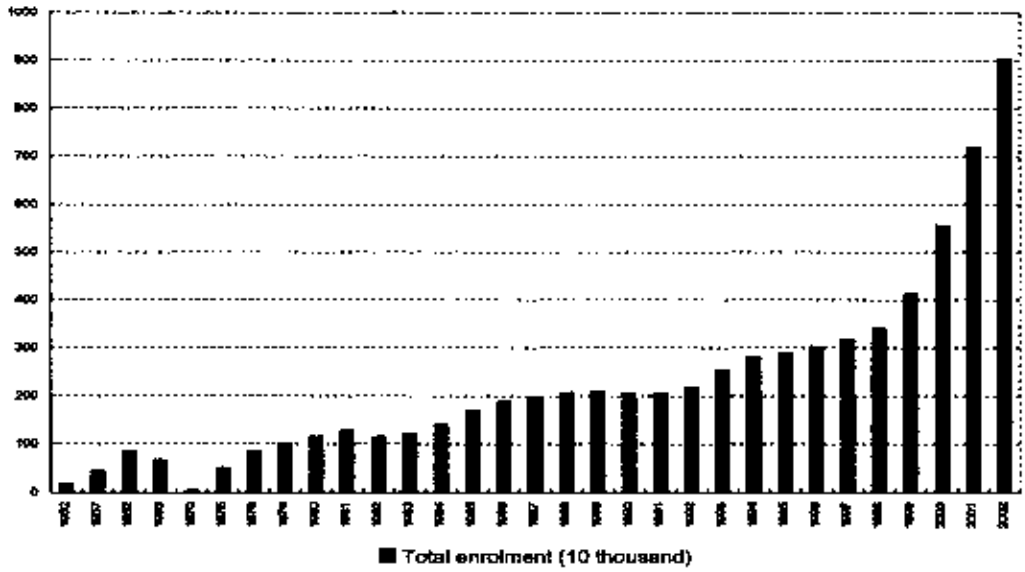
Figure 1.1 Basic Statistics of China's Higher Education (Public), 1980-2003



Source: National Bureau of Statistics of China. (2003) China Statistical Yearbook 2003, Vol. 22.

The number of public institutions granting college degree or above increased from 675 in 1980 to 1552 in 2003; total number of faculty increased from 0.247 million in 1980 to 7.25 million in 2003; new enrollment increased from 0.281 million in 1980 to 3.82 million in 2003; college graduate increased from 0.147 million in 1980 to 1.34 million in 2003.

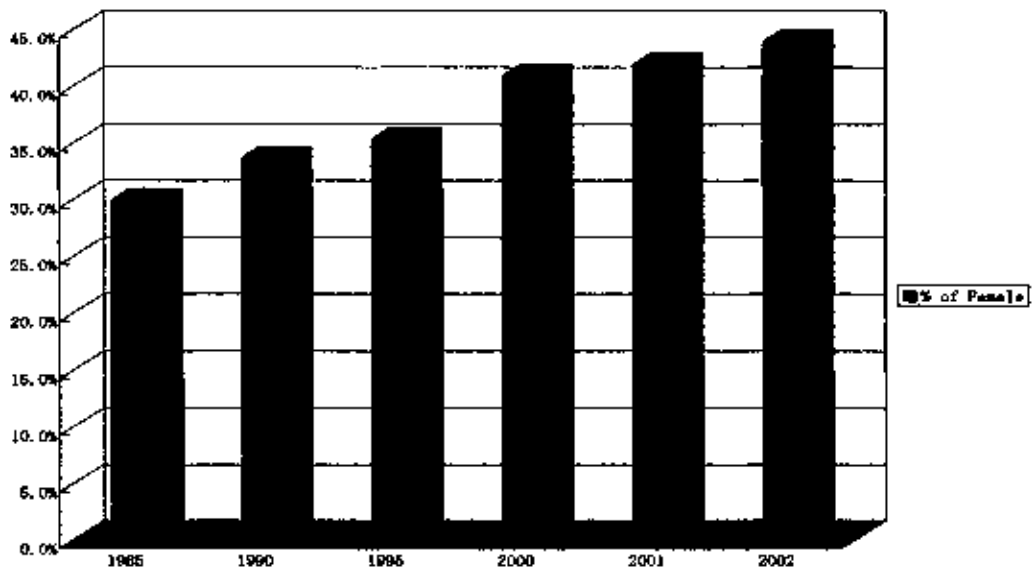
Figure 1.2 Total College Enrollments of Public Institutions, 1952-2002



Source: National Bureau of Statistics of China. (2003) China Statistical Yearbook 2003, Vol. 22.

Figure 1.2 shows the total public college enrollment from 1952 to 2002. The total college enrollment (undergraduate and up) reached 9.03 million, which was almost a 47 times increment from the 1952 level.

Figure 1.3 Female College Enrollment Percentages in Public Institutions, 1985-2002



Source: National Bureau of Statistics of China. (2003) China Statistical Yearbook 2003, Vol. 22.

As shown in Figure 1.3, by 2002, female college enrollment has been almost equal to the male counterpart and reached 44% after two decades' increase, which is approaching to the national population figure of 48.5% (Census 2000).

For more than half a century, from 1896 to 1949, Chinese higher education progressed according to the Western university model, although Chinese universities suffered heavily in the Resistance War against the Japanese Invasion (1937-45) and in the War of Liberation (1946-49). With the founding of the People's Republic of China in 1949, however, Chinese higher education cut off links to the Western world and turned, for various political reasons, toward the former Soviet Union's model for universities.

Educational policies in China in the past five decades have been characterized by bold moves, major shifts and reversals (Tsang, 2000). A reconstruction of Chinese higher education, involving reorganization of universities and disciplines, took place nationwide in the early 1950s (Duan, 2003). In this movement, comprehensive universities were reformed into single disciplinary universities or institutes such as institutes of engineering, medical colleges, agricultural colleges, broadcast institutes, and so on. Each reorganized university or institute offered many more majors in specific curricula than were available under the Western model. These far-reaching changes eliminated any real comprehensive university in China for nearly five decades. The current

reorganization of higher education, initiated in the late 1990s, involves a return to a truly comprehensive university (Duan, 2003).

Three waves of intense policy debate among Chinese leaders and educators on higher education influence the direction of its development in post-1949 China.

> "The first wave took place during the period of Great Leap Forward. As part of Mao's bold and ambitious national experiment with communism, the Chinese Communist Party (CCP) set a policy of rapid expansion and curriculum reform for higher education in 1958. The goal was that, in about 15 years' time, access to higher education would be provided to youth and adults from all backgrounds who qualified and were willing to attend higher education"... "The number of institutions increased from 791 in 1958 to 1289 in 1960. Student enrollment jumped from 660000 to 962000 in the two-year period, and increase of 46%. Obviously, such a rapid increase put a heavy burden on the financial resources of the government and had adverse effect on the quality of higher education. The failure of the great Leap Forward experiment led to the adjustment period of 1961-1965. " (Tsang, 2000, p. 608)
>
> "The second large swing in aggregate scale came during the period of GPCR (Great Proletariat Cultural Revolution)"... "Student enrollment as the higher education system was paralyzed by the political campaign of the time. Higher education admission was actually discontinued during the first haft of the GPCR. Thus, total enrollment declined from 534000 in 1966 to 48000 in 1970" (Tsang, 2000, pp609)

The third wave of higher education reform, which is the most recent policy adjustment, occurred in late 1970s. Since then, Chinese higher education enters a new ear. A detailed summary is provided in the following sections.

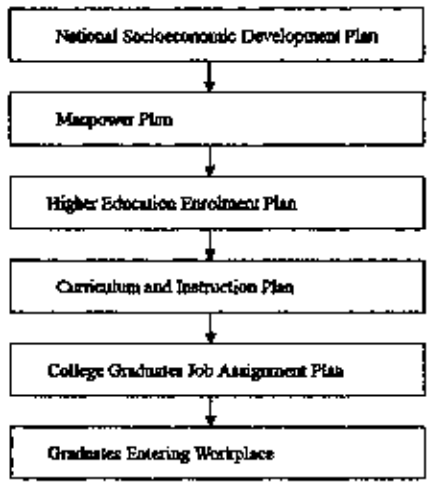
I.2.2 Higher Education Reform in China since 1980s

After the ten-year-long Cultural Revolution from 1966-76, the process of transition of the Chinese economy from the ossified centrally planned economy to a dynamic socialist market economy has led to a series of profound socioeconomic changes, which have strong impact on the Chinese higher education system (Min, 2002). As Chinese society changes, higher education is undergoing major transformations. The Western university model exercises a powerful influence, especially the American model.

The foremost issue for reforming Chinese higher education system is to establish a new institutional framework and operation mechanism fitting in with the new context of the dynamic market economy of the information age. However, the institutional framework and operation mechanism for Chinese higher education in the 1980s and 1990s was still basically the one which took its shape in the context of the centrally planned economy since 1950s. Higher education institutions formulated curriculum and instruction plans according to the specific national manpower requirements. Students were trained in a relatively narrow specialization. Graduates' job assignment plan was formulated by government according to the staffing plan of each line ministry and province. Figure 1.4 illustrates the structure of the old operation Framework of China's higher

education system.

Figure 1.4 The Operation Framework for Chinese Higher Education in the Centrally Planned Economy



Source: Min, 2002

The institutional reform included reorientation of the government—university relationship, stipulation of the legal status of higher education institutions, granting more autonomy to universities enabling them operate according to the needs of socioeconomic development and labor market demands, but not dictated by the government plan. The state will function through formulating higher education laws and providing policy guidance, and through coordination, evaluation and accreditation. Much has been done along this direction of reform in recent years, and the new operation mechanism is now gradually replacing the old one (Min, 2002).

In terms of finance, "free higher education" came to an end. For about four decades following the founding of the People's Republic, Chinese college students

did not pay tuition. If a student was admitted into a college or university, the student was guaranteed a "free" higher education. The central government subsidized students' tuition through their institutions. In the early 1990s, this situation was deemed incompatible with the growth of a market economy, and Chinese colleges and universities began to charge tuition and fees to students. At the beginning of the reform, students paid a small fee, but by the end of the 1990s, all college students had to pay their own tuition. In the meanwhile, more scholarships and grants have become available to help students, especially those from low-income and poor families in the countryside (Duan, 2003). Considering the affordability for the students, the Ministry of Education and provincial department of education cap the tuition and fee. Tuition plus living expenses represents a considerable amount of money for Chinese families, especially low-income ones.

I.2.3 Expansion of China's College Enrollment

The enrollment of higher education institutions in China rose from about 1 million in the early 1980s to about 10 million in 2001. In 1998, a massive expansion of higher education enrollment was launched by the Ministry of Education to reduce the gap between the strong demand for higher education and

the limited access to it. According to the official figures, the total enrollment for public colleges and universities increased from 6.43 million in 1998 to 19 million in 2003. The total of number of public institutions increased from 675 in 1980 to 2110 by the end of 2003 (MOE, 2003).

In 2001, the Ministry of Education abolished the limitations based on marital status and age for the national college entrance examination. Before this reform, candidates who were older than twenty-five or married could not take the examination. This significant change opened opportunities for adults to enter regular colleges and universities (Duan, 2003).

The higher education system expanded very fast, but the increase of state appropriation for higher education simply could not keep up with the growing costs, which led to a tight budget for universities. First, reform effort has been made to change the structure of government spending in favor of education. Despite the increase of the state appropriations to higher education since early 1980s, public expenditure on education in China remains relatively low by international standard. In late 1990s, China spent less than 3% of its GDP on education, as compared with an average of 6% for developed countries and 4% for other developing countries (Min, 2002). A decision was made in late 1990s that the rate of increase of appropriation to education at all governmental levels should be higher than the rate of increase of their revenue.

Along with the increase of willingness and capacity to pay, tuition level is gradually raised. At the present, about one fourth of total operational budget of Chinese public higher education institutions comes from tuition. Universities are also encouraged to improve their financial situations through enhancing management to improve institutional efficiency and effectiveness to turn a relatively high cost system into a more cost-efficient and cost-effective one. It could be achieved by internal reorganization of universities and colleges to rationalize small departments, broaden specialties, eliminate duplications of programs, and make more effective use of staff and physical resources including raising student—teacher ratio (from 3:1 in 1983 to 16:1 in 2000)and improving utilization of classrooms and laboratories (Min, 2002). One possible approach for cost-saving is to achieve economies of scale, which lies in consolidating small institutions into larger ones together with efficiency measures (Min, 2002).

In terms of governance, the Ministry of Education has played an active and decisive role in educational reform. One major change in governance has been the introduction of the "two-level education provision system," in which the Ministry of Education shares responsibility for educational governance with provincial department of education. The provincial department of education has been assigned greater responsibilities and now directly administers most regular universities and colleges (Duan, 2003).

I.2.4 The Emergence of Non-government Tertiary Education

The non-government tertiary education includes two sections: (1) *Minban* (non-public) post-secondary education and, (2) out-of-plan-fee-paying college enrollment within public institutions.

Since mid 1980s, the government has encouraged organizations and individuals to found private post-secondary institutions. These colleges are so called *Minban*, or "people run" school. There are three levels of *Minban* institutions, which are differentiated by the ability to confer degrees and diplomas. Type 1 institutions can confer their own degrees. Type 2 institutions can grant joint diplomas. Type 3 institutions prepare students for self-study national examinations. The latest statistics indicated the there were 1282 *Minban* tertiary institutions by 2000 (MOE, 2000). However, only 24 of these colleges are accredited by the Ministry of Education and are able to issue bachelor degree.

Students admitted to *Minban* institutions are totally outside of the national undergraduate recruiting plan made by Ministry of Education. Some observers estimate that there are currently 4 million students studying at privately funded tertiary institutions in China by 2000 (LaRocque & Jacobsen, 2000). Unfortunately, the real enrollment of *Minban* sector is unavailable so far. The newly passed

Minban Education Promotion Law by the 9th Committee of the Chinese National People's Congress legalizes the status of *Minban* Institutions.

Apart from *Minban* Institutions, another section of non-government college education is out-of-plan-fee-paying enrollment in the public institutions. Though these students are admitted to the public institutions, their matriculations do not occupy the national plan quota of their institutions designated by the Ministry of Education. They pay higher fees and tuition to the public institution than the "normal" students, who are admitted under the national plan. These out-of-plan students, in general, have lower College Entrance Examination (CEE) score than their "normal" matriculation colleagues in the public school, but perform better academically than students in *Minban* schools. Public institutions recruit more students out side of the national plan in order to pursue larger marginal productivity of the education resource and, increase the revenue to cover the deficit by the gradually reduced government appropriation.

I.2.5 College Education and Job Placement

Before the 1980s reform, in return for a free higher education, Chinese university students were to take whatever jobs the government assigned to them upon graduation. Graduates could not choose their place of work or their jobs.

After several years of reform, the centralized allot system was abandoned by the central government. All Chinese college graduates must now find their own employment. In the old system, all graduates received the same starting salary no matter what they had majored in, or what kind of work the government assigned to them. Now, wages of new graduates vary depending on the classification of their job, the region in which they work, and the offerings of individual employers.

The structure of Chinese higher education system in the centrally planned economy was characterized by departmentalization and segmentation. 358 national level universities and colleges were under 62 different ministries (Min, 2002). After the higher education reform, however, when a student graduated from a university belonging to a specific ministry, he or she might find a job in another trade through labor market mechanism. Thus the manpower training plan of original ministry specific system failed. The older allot manpower system does not fit the new higher education system anymore. Consequently, the manpower plan was abolished. To fulfill the demand of labor market, the Ministry of Education restructured Chinese higher education system through mergers of universities or setting up collaborative arrangements among higher education institutions by breaking the former departmentalized boundaries of different ministries. Nationally, there were 612 universities and colleges merged into 250. The structure of Chinese higher education system was changed dramatically (Min, 2002).

I.3 Definition of Terms

Socioeconomic Status (SES) is a term that is used to describe factors affecting a person's lifestyle including occupation, income, and education. Hauser (1975) put that socioeconomic status includes characteristics as parents' education, occupation, and income; racial, ethnic, and religious origins; the environments and opportunities offered by home, community and school; and individual characteristics like abilities, aspirations, and efforts. SES influences the education attainment, occupational achievement and earnings of an individual. The concept of SES is widely adopted by social science researches such as in economics, sociology, political science, education and etc. This study adopts the concept of socioeconomic status mainly from the economic perspective to examine the correlation between college graduates' socioeconomic background and their initial earnings.

Initial Earnings in this study refers to the initial salary earned by the college graduate from his or her first job offer after graduation. For the sake of computational convenience, all initial earnings are expressed in Yuan /year/person.

Employment Status or **Current Employment Status** in this study refers

to whether a college graduate gets a job offer by the end of the survey, i.e., June 20th, 2003. If by that time, a college graduate had a job offer, then his or her employment status is "employed", otherwise "unemployed".

Matriculation Type refers to the enrollment status of the college students in this study. More specifically, there are two matriculation types in this study: matriculation with national plan and, matriculation without national plan. The national plan is the recruitment quota designated to each individual public tertiary institution by the Ministry of Education. The ministry makes national enrollment plan every year, and distributes the quota among all public institutions. Before the latest higher education reform the national plan, or quota was very rigid since the quota directly link to the appropriation to the institutions. Public institutions were fully funded by the government at that time and the students got free higher education. Therefore, colleges and universities could only recruit the designated number of freshman every year according to the quota. However, after the higher education reform from the middle of 1980s, alone with the marketization of higher education and labor market, the financial source of the public institutions were greatly diversified. And the government appropriation as percentage of the total school revenue declined significantly. To encourage the public institution fully utilize the existing resources to support themselves and educate more students, the Ministry of Education loosened the restriction on the national plan and allowed the

public institutions to recruit students out side of the quota with slightly lower College Entrance Examination (CEE) score. As a condition, the public institutions can charge higher tuition and fees from those out-of-plan students.

College Entrance Examination (CEE) was and still is the only measure of college admissions. High school graduates have to attend the CEE test and apply for the colleges and universities according to their CEE score. Different levels of institutions recruit new students based solely upon the CEE score, only high scoring students can go to prestigious institutions. The quality and prestige of schools drop with the candidates' CEE score accordingly. Due to the scarcity of the supply of higher education, admission of public institutions is highly competitive. Those low CEE score students might have to take CEE test again and again. Since the reform in the mid-1980s, the Ministry of Education loosened the national plan, which also eased the restrictiveness of the college admissions. The MOE began to diversify the type of institutions and encourage the private sector to run the non-government post-secondary Institutions in order to expand the college enrollment. A lot of so called *Minban* institutions were founded even since.

Ownership of Institutions in this study has two types: public and non-public. The fundamental differences attribute to what kind of entity fund and run the institutions. Before the reform, all institutions are public. However, after the reform, individuals and private entities were allowed and in fact encouraged to

establish and run non-public institutions. In a certain occasions, public entities and private entities jointly found institutions. These hybrid ownership institutions together with the pure private ownership institutions are defined as *Minban* institutions in China. The term of *Minban* is also adopted by this study.

Origin of the students in this study refers to the original residency of the students. In China, domestic migration was and still is restricted. People are not allowed to move from place to place freely in the central planning economic era. City residency, county residency and rural residency are separated strictly. The social welfare, access to social resources, and the living conditions of these three origin types differentiate greatly. This hierarchical system is called *Hukou* system. Though the *Hukou* system has been widely criticized by the public and researchers for years and has been loosened gradually after the Open Policy, the origin of an individual still affects largely his or her social economic status nowadays. The *Hukou* system is not abandoned yet.

Higher education experience refers to the academic achievements the students make during the entire college years. In this study, several indicators help to present the higher education experience such as the overall class ranking, College English Test (CET) passage and merit-based scholarship awarded.

CHAPTER II

LITERATURE REVIEW

II.1 Theoretical Perspectives

This chapter discusses the theoretical perspectives for understanding individual's decision to invest in higher education based on their socioeconomic status and projected future earnings. A concise review of the literatures on the extent of college choice and employment choice in China and the United States is presented. A review of the literature conducted in the American context is provided primarily for the richness of the data available and highly developed debate about explanations of earnings differences.

Decisions to invest in post-secondary education or to enter the labor market after high school graduation are not spontaneous. High school graduates can enter the job market and earn money, yet a lot of them choose to go to college. Given that attending a higher education institution is not only costly but also time consuming, why do people still invest in higher education? The response to this

question can be based on several theoretical perspectives, including human capital theory, status attainment model and social conflict perspective.

Though students and their parents might or might not know these theories, their decisions are still consistent with the principles of these theories. Intuitively, everyone has the propensity to move upward to a higher socioeconomic segment. By increasing the social mobility, one is able to achieve a higher socioeconomic status. An increase of personal wealth is a major means to gain higher socioeconomic status; therefore, people always seek to increase the present value of his or her total wealth (Mincer, 1974; Rumberger, 1983). For most of the people, life-long earnings are the major piece of their wealth, if he or she doesn't have a large amount of inheritance from the parents. People tend to optimize their earnings, according to the golden role of optimization in the mainstream economics. They tend to estimate the future earnings and utilities of education investment, calculate the monetary cost and opportunity costs of schooling, and select the best choice of working or schooling (Hanushek & Quigley, 1985).

II.1.1 Human Capital Theory

Human Capital Theory, first introduced by Mincer (1958) and Schultz (1961) and further developed by Becker (1964), Denison (1962), and others claims

that individuals and society derive economic benefits from investment in people. For individuals, the skills, knowledge, and dispositions acquired through education raise the productivity capacity of an individual. A more productive individual is rewarded with a higher wage in a competitive labor market. Education is no longer viewed as a kind of consumption good but as an investment that improves their future earnings and subsequently the range of choices available to the investors.

Although types of human capital investment generally include health and nutrition (Schultz, 1981), education consistently emerges as the prime human capital investment for empirical analysis. Schultz (1963), points out that education is perceived to contribute to health as well as nutritional improvements, and education may be measured in a quantitative manner (dollar costs and years of schooling) in the empirical studies. Human Capital literatures distinguish several types and means of education as human capital investments, such as formal education at primary, secondary and higher levels (Cohn & Geske, 1990), informal education at home and at work (Schultz, 1981), on-the-job training and apprenticeships (Mincer, 1974), and specialized vocational training (Corazzini, 1967).

Human capital investment contributes to national development and economic growth too. A lot of researchers studied education's contribution to improvement of population control, overall quality of life, civil society, social

participation, democracy, national productivity and so on (Swanson & King, 1991). All these social benefits of human capital investment are usually called externalities. However, most of these externalities are difficult to quantify.

Becker (1993) provides an extensive economic analysis to explain how individuals make decisions regarding higher education. Investment in schooling, a formal form of education, can be understood from a set of supply-demand curves where the supply of investment in schooling refers to the capital required to finance this investment and the demand refers to the need of the labor market. Subsequently, people with more favorable social opportunities tend to invest in more schooling and people with higher abilities and of younger age tend to invest more as they have a greater capacity to benefit from educational investments. Since both demand and supply curves differ from individual to individual, different people could invest in the same amount of education and yet some could earn more than others. Additionally, demand and supply curves are likely to be correlated. For instance, a more able person is more likely to receive a merit-based scholarship, holding others equal. Thus, distribution of earnings depends on the distributions and shapes of these supply and demand curves. Becker concludes that earnings are more unequally distributed and skewed the more responsive the supply and demand curves. Becker argues that greater equality of opportunity (supply) not only tends to reduce the inequality in earnings but also increase the

efficiency of allocation.

Becker's analysis explains differences in people's investment in higher education. Considering one's decision to invest in higher education as a cost-benefit analysis where one will invest in higher education only when its perceived earnings outcome outweighs the cost of the investment. Thus, according to this framework, schooling costs and financial aid policies and one's ability will influence one's decision of higher education investment, and then consequently influence the return to higher education.

Human Capital Theory has also been criticized on several grounds. Fagerlind and Saha (1989) note that, at the individual level, it is highly controversial whether human capital investments are directly linked to improvements in productivity or income. Other factors like family background, personal ability, or the prevailing social, economic and political systems, may better explain individual success. Solow (1965) argues that when calculating the returns to human capital investment, several types of selectivity bias are very easy to mislead the results and interpretation underlined. At the societal level, the theory is problematic as it attributes the unexplained economic growth primarily to social investment, the most of which is education.

II.1.2 Social Conflict Perspective

The social conflict perspective of Carnoy and Levin (1985) stresses the contradictory role of schools to reproduce the existing hierarchical relations of the capitalist workplace, which is usually unequal, and to present opportunities for social mobility and the extension of democratic rights as well. Carnoy and Levin also assert that the dynamics of the US educational system can best be understood as part of a much wider social conflict arising in the nature of capitalist production with its inequalities of income and power and these inequalities lead to struggles by subordinate groups for greater equality, economic security, and social control. They indicate that in a politically democratic society, the democratic State provides space for such struggles. The educational system, being situated within the democratic State, becomes part of such conflict. Which of these movements dominates is determined by the larger social conflict and the relative political strength of the groups involved. The struggle of dominated groups to change the conditions that oppress them and the attempts of dominant groups to reproduce the conditions of their dominance are the key to understanding changes in the economy, in social relations, and in education. These changes, in turn, are reflected in State policies and in public schooling.

The social conflict perspective helps to identify the reproduction nature of

education processes and provide an understanding of how students of different SES groups are channeled into different curricular tracks, how students are socialized into different educational goals through parental expectations based on their SES, how students of different backgrounds learn different things and have varying academic preparation, and how the schooling process reinforces social-class distinctions and socioeconomic outcomes after schooling.

II.1.3 Status Attainment Model

As formulated by Duncan and Featherman, (1972), status attainment model assumes that social status of parents affects educational level achieved by children, which in turn affects occupational level and status (Hotchkiss & Borow, 1996). The level of schooling would affect or moderate the degree of intergenerational transmission of social status. One may incorporate a mental ability variable into this model, assuming that it also predicts educational achievement (Sewell & Hauser, 1975). Sewell and Hauser (1975) point out that a number of experiences that young people undergo in their formative years have an important bearing on post-secondary educational outcome, including higher school performance, encouragement or discouragement of higher education, occupational aspirations. There experiences intervene between social origins, academic ability, and sex

characteristics of the individual and his or her later achievements and become the mechanism through which there background characteristics transmit their influence. Moreover, these same social psychological experiences have direct and indirect effects of their own, quite independent of the youth's background characteristics (Sewell & Hauser, 1975).

Sewell and Hauser's social attainment model links socioeconomic status and academic ability with educational and occupational attainment by means of social psychological variables such as academic performance and aspirations. Their study finds evidence that parents' income, education and occupation significantly influenced the children's status attainment (Sewell & Hauser, 1975). Individual, family and neighborhood characteristics are all found to have large and significant effect on the probability that a student will repeat a grade in Corman's (2003) study. Alexander and Eckland (1977) argued that social status composition of the high school was found to enhance student's prospects for attending a selective institution of higher education, yet the effect was significant for male only. College selectivity, in turn, had total salutary effects on educational attainment, despite its depressant effect on undergraduate grade performance and academic self-concept. Sewell and Hauser (1975) also found that the effect of socioeconomic origins and academic ability on educational achievements and occupational attainments as these influences are mediated by social psychological

processes. Arthur J. Corazzini and et al. also found strong structural relationships between college attendance and socioeconomic status (Corazzini, Dugan & Grabowski, 1972). Smith and Powell (1990) found that college education gave students reasonable understanding of the value of a college education comparing to the high schooling graduates, however, men's self-enhancement of aspiration are stronger then woman and more prestigious institutions influence students more on income aspiration.

Paul Lindsay and William Knox's (1984) research found that educational attainment affects both intrinsic and extrinsic work values significantly. The more education, the more likely people are to value the intrinsic rewards of work and the less likely they are to value the extrinsic ones. Educational attainment not only influences work values but also affects job characteristics. Higher education operates to allocate people into occupations with more self-direction and ideational content. The more education the greater the probability of extrinsically rewarding work, for education is related to earnings. The socioeconomic status is a significant predictor of educational attainment. Education has both socializing and occupational-selection consequences, and educational selection is itself a consequence of previously existing work values. Educational attainment is a key placement mechanism in occupational selection. Rytina (1992) found that offspring of the most favored occupations are the most successful in obtaining

education, but the continuation of occupation across generations is substantially independent of the educational accomplishment of offspring.

Since parental socioeconomic status influences the offspring's schooling choice and education attainment, and education attainment closely relates to one's occupational attainment, an intergeneration cycle is formed. The cycle could be both virtuous and vicious. That is, on the one hand, the influence could be positive to the high SES families, yet on the other hand, it could be negative to low SES families. Status attainment model reveals this intergeneration relationship and numerous related literatures provide solid empirical evidence.

II.1.4 Heckman Two-stage Method

James Heckman's (Heckman, 1974; Heckman 1979) two-stage method offers a computationally simple procedure for handling selection bias when data are generated by a non-random selection process. Selection bias arises when a sample under study does not randomly represent the underlying population. The problem facing the analyst is to obtain estimates of relevant population parameters even in the wake of a selective sample. Non-random sample selection may result from individual decisions by the agents under study (self-selection), but may also reflect administrative rules, or decisions on the part of sampling statisticians.

Selection problem can be viewed as a problem of missing observations. For instance, there is a lack of information on the earnings of workers with a high-school education, had they pursued a university education. Heckman's key insight is that observations are often missing because of conscious (self-selection) choices made by agents (e.g., the decision to work, to pursue higher education and etc.). Heckman's contributions to the econometrics of selective samples emerged concurrently with his Heckman Two-stage method, or Heckit method in the mid-1970s, which was further developed by Willis & Rosen (Willis & Rosen, 1979), Lee & Robert (Robert & Lee, 1984) and others, which are distinguished by estimating equations derived explicitly from utility maximization with stochastic error terms as an integral part of the model, rather than added as an afterthought. They enabled a unified analysis of the factors determining work earnings and labor-force participation.

II.2 Empirical Studies

The following is a brief review of literatures on how socioeconomic status influences the early earnings of college graduates and how education, especially higher education, is affected by socioeconomic status (SES). SES is a term that is

used to describe factors that affect a person's lifestyle including occupation, income, and education. It is typically used as a shorthand expression for variables that characterize the placement of persons, families, households, census tracts, or other aggregates with respect to the capacity to create or consume goods that are valued in the society (Hauser & Warren, 1997).

II.2.1 SES and Earnings of College Graduates in the United States

The benefit and cost of investing in higher education can be understood to play a critical role in one's decision to attend an institution in higher education in American. Henretta and Campbell's (1978) study found that the effects of family background are transmitted via education; the effect of education is asymptotic rather than linear; single and divorced persons possess substantially fewer assets, net of other characteristics, than married persons; and net of all other variables, earnings have a substantial effect on net worth. The effects of family background and socioeconomic attainments on net worth may yield a more stratified system. Hauser and Warren (1997) also found the positive intergenerational relationship between occupational education and occupational earnings and the levels of occupational education appear to define the main dimension of occupational persistence across and within generations. In addition, Sewell (Sewell & Shah,

1968) and others (Hauser & Sewell, 1986; Sewell, Haller & Portes, 1969; Murphy & Welch, 1989) found that SES elements like race, parents' education, gender, parents' encouragement all accentuate the differences in college choice, earnings, and even SES itself in certain weights.

Lazear (1976) used the National Longitudinal Survey (NLS) data from1966 to 1969 and estimated that about one third of the total compensation of young workers attributes to human capital investment.

Murphy and Welch (1989) documented changes in the economic return to college observed over two decades. Overall returns to college education increased from 47% in 1963 to 61% in 1971, yet the returns declined from 61% in 1971 to 48% in 1979; and for young workers, Internal Rate of Return (IRR) to college education rose from 41% to 44% and dropped to 32% in the same period. They also found a dramatic change of returns to college education from 1979 through 1986, when returns rose from greater than average growth in the college population during the early and mid-1970s.

> "This greater than average growth in supply exceeded the rate of growth in the demand for college trained workers and depressed the wage premium of college graduates in general and young college graduates in particular. In contrast, the rapid growth in the college wage premium in recent years (1989s) seems to be result of a slow-down in growth of the college population and greater than average growth in the demand for college educated workers." (Murphy & Welch, 1989)

However, Murphy and Welch were not able to identify the precise cause of

the rise and decline of the returns.

Heckman (1979) introduced his two-stage method to the field of education—earnings research and developed the tractable technique to use non-randomly selected samples to estimate behavioral relationship between education and earnings. The method was soon adopted by several researchers in their studies of American labor market and higher education. For example, Willis and Rosen (1979) use Heckman's model and NBER-Thorndike-Hagen survey data of 1968-71 estimate that expected lifetime earnings gains influence the decision to attend college. They also find that those Americans who did not attend college would have earned less than measurably similar people who did attend. Hauser and Sewell (1986) reexamined the Wisconsin Longitudinal Study data with the Heckman two-stage method and found that family background have large independent effects on students' ability, schooling, and to a lesser degree, socioeconomic attainment. Kenny and et al. (1979) also found in their study that both self-selection bias and simultaneity bias are negligible for the college education group whereas there is some self-selection bias for the groups with no college education and the coefficients of the explanatory variables in the wage equation are significantly different form the two groups.

Race and gender have been consistently found to be significant determinants of earnings and higher education experience. Wanner and Lewis

(1982) examine the U.S. Census data from 1950 to 1970 and find that overall inequality of earnings is stable over the three decades. Furthermore, they also find sizeable increases in inequality of earnings among some lower SES occupations and at the same time decreases in inequality of earnings among higher SES occupations. Though education's effect on relative level of earnings shows no consistent pattern overtime, its effect on inequality of earnings becomes stronger overtime. Their study also finds the inequality of earnings for nonwhite worker declines between 1950 and 1970. Wanner and Lewis's findings is concurred by Haider's (2001) study. Haider analyses of the 1968-92 panel data and claims that earnings inequality for males in the United States increased during the early 1980s and the increase trend is related to changes in the returns to education. Cotter and et al. (1999) also find that the gender differences across racial/ethnic groups and at each earnings level are quite substantial. The permeability of racial and earnings boundaries to gender dynamics is quite impressive and gender boundaries are quite permeable to racial inequality. Hauser's recent study (Robert; Sheridan; Hauser, 2002) confirms the previous finding that differences between men and women in career trajectories and returns to schooling are substantial.

II.2.2 SES and Earnings of College Graduates in China

In contrast to the studies of the interaction between SES and earnings of college graduates in the United States, fewer quantitative studies were conducted on this topic in China. Though education, especially higher education, has been placed high value in China's Confucian tradition and culture, the rate of return to higher education had been remained at a very low level in the central planned economic era.

In recent years, in the economic reform period, Johnson and Chow (1997) and Liu (1998) used the 1988 Chinese Household Income Project data set and estimated that rate of return to higher education is only about 3 to 4 percent. However, Li (2003) used the same data set yielded different estimates. Li claimed that the overall rate of return to higher education in China should be around 5.4%; 4.3% for man and 6.9% for women. Li also argued that the lower previous estimates were due to the methodology problem, e.g., hourly wage should be used to estimate the internal rate of return instead of monthly or annual earnings. Li's argument is that in China, the most educated tend to work for the fewest hours on average (Li, 2003). He even estimates that college graduates work 3 hours less then middle school graduates (Li, 2003). Even the 5.4% returns to higher education is still much lower than the 14.4% level of the other less-developed countries (Liu, 1998).

Another reason of such a low returns to higher education is the non-satisfaction of the basic assumption of Human Capital Theory. So far, all the rate of return computation is based on Human Capital Theory model, and the theory has a fundamental assumption, which assumes the individual will decide how much higher education to obtain and what type of work to undertake in the free choice of the labor market. However, this assumption is not true for those studies using data earlier than 1990s. Because at that time, higher education was not accessible to everyone, and job assignments were largely determined by the central planned allot system instead of the labor market. This situation was not changed until the mid-1990s. Therefore, all rate of return studies in China before mid-1990s based on the human capital framework were in fact problematic.

Liu (1998) argues that economic and higher education reform after the 1990s brought significant changes in the employment system. The reform aimed to better link wage levels with both individual and enterprise performance so that the wage system could be used as a price mechanism to regulate labor movement. In addition, a growing non-state sector has created employment opportunities outside state-owned enterprises. Non-state-owned firms, especially foreign-invested enterprises, have tried actively to recruit highly skilled personnel by offering higher wages and salaries comparing to the state-owned ones. Consequently, state-owned firms are under pressure to increase their pay to maintain and recruit

high skilled worker and personnel. Therefore, the former central planned job allot system and compensation system became obsolete, and were changed greatly to match the market oriented labor market and national economy (Liu, 1998). And only after the abandoned of the old job placement and compensation system, the human capital model rate of return computation becomes meaningful. According to Liu's (1998) estimates, rate of returns to university education is 18.4%, to secondary education is 11.6%, and to primary education is 7.5%. His study also found that the average earnings of the working individuals in more developed coastal provinces are 1.54 times as high as those in other provinces, and rate of returns to younger workers is higher than that of senior workers, implying that economic reform benefited younger workers more than older workers (Liu, 1998).

Wu's study (2002) using two urban survey data, reveals that while the influence of redistributive power declines with the reform of the former central planned wage system, returns to human capital do not monotonically increase as a firms' proximity to the market increases. Returns to human capital are higher in the market sector than in the state sector, the effects of education on earnings are weaker in the high-profit state-owned firms than low-profit state-owned firms within the state sector. The inconsistency is attributed to the effects of bonuses that are equally distributed among employees in high-profit state-owed firms.

Bian and Logan (1996) documented that during the piecemeal reform

period in the 1980s urban income equality was improved by the salary increases of low end workers. However, during the more extensive reform period in the 1990s, incomes became more unequal due to further marketization of the labor market and national economy.

As to the relationship between higher education and employment, recent research conducted by Zheng (2004), using a survey to college graduates of year 2002, found that family SES has an immediate effect of various degrees upon graduates' employment will, job-application behavior and the possibility of success. Higher SES graduates are more likely to suspend their employment, or to choose high pay occupations, and to have higher income expectation.

Wei and et al. (2000) found in their study that there existed a great difference of the starting salary of college graduates from different level of institutions. The starting salary of key university graduates was 1.5 times of those graduated from ordinary colleges, and the difference of annual increase of salary between two groups could be as large as 21.83 times (Wei and et al, in Tsang, 2000), though in the study, the authors didn't estimate the influence of family socioeconomic status' affect on starting salary of college graduates.

Another recent study (Zhao, 2000) reveals that the differences of family SES among male college students are larger than female ones, and female students' family SES has stronger correlation with their opportunity of having

higher education. Yu and Lu's (Yu & Lu, 2001) study find out that higher education enrollment helps to mitigate the household socioeconomic disparities. That is, the college enrollment ratio of 6-year old and up age cohort is negatively correlated with the local Gini coefficient. Xun's study (Xun, 2004) argues that the changing of the allot system in China causes large disparities in the income distribution, which lead to the unequal starting points of employment.

Li and Wang's (2004) recent study found that substantial sorting gains under the traditional (job placement) system but that gains have diminished and even become negative as higher education choices widened and participation has become subject to increasing direct private costs. They considered this as evidence consistent with the influence of financial constraints on decisions to attend college.

So far, most of the studies in China generally use simple statistical correlation to examine relationships among variables. The analysis on employment and earnings do not correct for self-selection bias. Only latest study done by Fleisher and et al. (Dec., 2004) applies Heckit model to examine the self-selection bias, however this study focuses on sorting gains of the college education and does not examine the family SES's influence on college graduates' initial earnings.

II.3 Gaps in Knowledge

The literature review from the United States highlights the importance of having detailed student-level data so that research analyses on the relationships among SES, higher education experience and earnings can be better understood. The smaller literature in China on this topic is mainly due to the lack of comprehensive census data that include students' family SES backgrounds and rigorous quantitative analysis. The available literature in China suggests that student socioeconomic backgrounds do have impact on their higher education experience, and earnings. The few available studies only used simple correlation comparisons or descriptive statistics rather than using econometric analyses to separate out the effect of specific factors while keeping all other factors constant. Most of the studies do not apply advanced econometric methods to correct the selectivity bias and the results are problematic due to the truncate selection problem. Some early studies even have problematic assumption in the Human Capital Theory framework and the Mincer formula.

There is no nationwide study focusing on family socioeconomic status' interaction with higher education experience to identify detailed determinants of the initial earnings of the college graduates in China. For a fast growing economy like China, such study will provide valuable information to educational policy-making and future socioeconomic structural adjustment.

This study proposes to use advanced econometric methods, with truncate selection bias controlled, to understand the current situation of the relationship between family socioeconomic status, college education and initial earnings in China.

CHAPTER III

KEY RESEASRCH QUESTIONS, ANALYTICAL FRAMEWORK, AND RESEARCH METHOD

III.1 Key Research Questions

The purpose of this study is to focus on understanding the interaction among family socioeconomic status, higher education experience and college students' initial earnings in China. Adopting the Human Capital Theory and status attainment model and facilitated by Heckman two-stage estimation method, this study attempt to address the following four key research questions:

1. Does higher education experience affect college graduates' initial earnings?

2. Is there any significant difference in initial earnings for college students from different socioeconomic backgrounds? How different are they?

3. Does the effect of family socioeconomic status on college graduates' initial earnings operate through higher education?

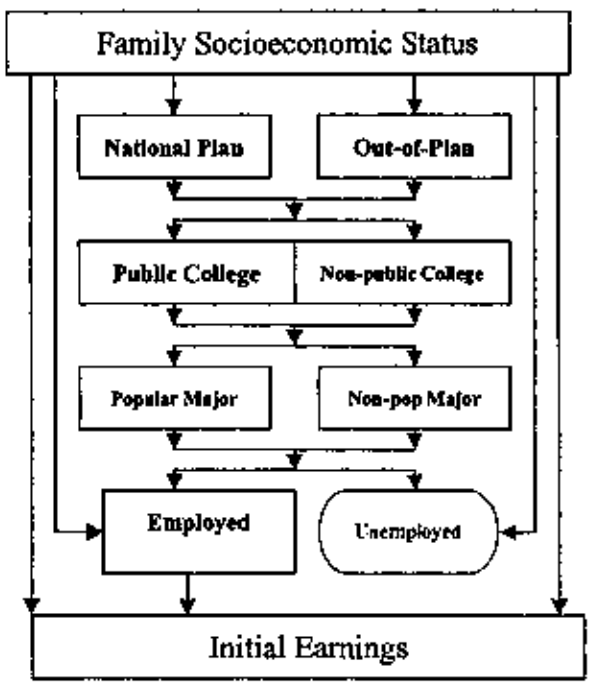
4. Are factors that influence initial earnings different for male graduates and

female graduates?

III.2 Analytical Framework

A SES-Higher Education-Earnings interaction model could be presented in the following:

Figure 3.1 SES-Higher Education-Earnings Interaction Model



For a student, his or her family socioeconomic status takes effect even at the very beginning of the higher education streaming and in turn influence the future earnings after the graduation. The model indicates that family background exerts a direct influence on adult economic status in addition to its indirect influence on

education (Rumberger, 1983). Going to a more prestigious public college or a lower level non-public college is influenced by his or her family SES background. Even the student does go to a public institution, whether he or she can get admitted under the national plan or admitted with an out-of-plan matriculation status is also influenced by the family socioeconomic status (Liu, 2003). Students are streamed into different type of institutions with different majors in college. Though individual effort has impact on higher education experience, the initial and persistent impact on family socioeconomic status has already let to an unequal starting point to college education. Eventually, all these endogenous and exogenous factors work together to determine students' initial earnings when they graduate from college.

The model reveals that family socioeconomic status has significant impact on students' higher education experience and their initial earnings, which in turn will later have impact on the students' own SES. The relationship among SES, higher education and earnings becomes a circle—virtuous or vicious, which will affect a person's social attainment on the one hand, and will transmit from generation to generation on the other hand. Eventually, one's family socioeconomic status will not only affect his or her own educational attainment in his or her own lifespan, but also influence his or her next generation's social attainment and mobility. This intergenerational relationship is not examined in this study.

As mentioned in the previous literature review, sound quantitative empirical

studies on this topic is sorely lacking in China. In order to fill in the knowledge gap, this study focuses on the latent relationship between family socioeconomic status and graduates' initial earnings through higher education. Several models and theories are therefore, adopted.

The model used in this investigation is based on two theoretical foundations. The first is the Human Capital Theory. As Schultz (1961) mentioned in his famous *Investment in Human Capital* that education is highly germane to human capital investment. Specific to this study, college education is treated as a form of human capital investment, which is believed to have impact on their earnings. The students' different selections of college education, such as tastes of college education, characteristics of institution and different majors are introduced as predictors of the earnings equation to estimate the future earnings. Human capital investments like investment in health and migration of individuals and families to adjust to changing job opportunities are omitted since the data are not available with the current data set. Other human capital investments such as on-the-job training and continuing education are not applicable for the targeting group—college graduates, therefore are also omitted. However, Human Capital Theory alone is not sufficient to capture the characteristics of return to college education and has been criticized on several grounds. At the individual level, it is highly controversial whether human capital investments are directly linked to improvements in individual productivity or

earnings. Other factors like family background, personal ability, or the prevailing social, economic and political systems, may better explain individual success. Becker himself (1966) also further adjusted his Human Capital Theory to involve family background as a factor. Hence, the educational / social attainment model is adopted as the second theoretical foundation of this study.

According to the status attainment model, college characteristics and family socioeconomic status are viewed as inputs in the production of the students' status attainment. Specific to this study, the output is the earnings of the college graduates; school factors include type of institutions, ownership of the school, college majors and matriculation status. Family socioeconomic status refer to parents' occupation level, parents' education level, family financial assistance, student loan, and miscellaneous expenses. The inputs of these school and family attributes are believe to the impact the output.

Both of the preceding theories are constructed in a way such that individuals make decisions about higher education based on their family socioeconomic status and eventually influence individual's status attainment via education attainment both directly and indirectly (Rumberger, 1983; Grilches, 1977). On the one hand, SES factors affect the college education process first, and then indirectly affect the human capital investment outcome, i.e., earnings. On the other hand, the family SES factors, especially those latent family socioeconomic status variables, affect the earnings

directly. It is difficult to differentiate between the direct and indirect impacts, since in some occasions, family SES factors have impact on both higher education experience and status attainment, which means their impact are both direct and indirect. In this study, the approach to obtain the influence of the family socioeconomic status to the college graduates' earnings is designed to be indirect: that is, to first construct an earnings equation without controlling the impact of the family socioeconomic status, and then construct a second earnings equation with family SES controlled; and finally compare the differences between these two earnings equations. Since the difference between the two earnings equations are the family socioeconomic status, the differences represent the impacts of these SES factors, no matter they are direct or indirect.

The first earnings equation without socioeconomic status control is a linear multiple regression equation. The independent variables represent higher education experience factors and other control variables. The second earnings equation is a Heckman two-stage method equation with family socioeconomic status controlled. The merit of the Heckman two-stage method is to control the self-selection bias. The core hypothesis proposed in this investigation is: family socioeconomic status (SES) impacts the college graduates' initial earnings through college education. Under this hypothesis, there are two sub-hypothesis, or say, two scenarios under the principle hypothesis, i.e.: (1) some family socioeconomic status factors impact the earnings

directly through college education; (2) other SES factors impact the earnings indirectly by affecting the selection or choice of education and employment.

Under the analytical framework of this study, direct socioeconomic status factors are responsible for earning variations in the first scenario, while earning gaps in the second scenario attribute to the indirect SES factors, or latent SES variables. For scenario one, direct SES factors refers to those independent variables, which are germane to family socioeconomic status and impact individuals' earnings in a measurable manner. For instance, miscellaneous expenses, which is the aggregated expenditures of everything besides tuition, fees, lodging, food expenses transportation cost and job seeking cost, reflects the largely the specific student's family wealth and personal financial situation. It is reasonable that a wealthier college student spends more in general than a student from a low SES family. The impact of direct SES factors on the specific student is direct and measurable, e.g., one's miscellaneous expenses can be measured by the amount of money he or she spends and doesn't require any medium to take the effect on that student. Latent SES factors, on the contrary, can either not be measured in an accurate manner, or not be able to impact the subject directly without a medium. For example, parents' occupation level, which is not a scale variable (ordinal at the most) and does not impact the subject—the student directly. A medium is required to transmit the effect, e.g., parents' occupation impact the student through his or her parents. Therefore, it

is very difficult to measure or even identify the impact of the latent family socioeconomic status to the student's earnings.

The unique design of the conceptual framework of this study is to utilize Heckman two-stage method to capture the latent SES factors' impacts and improve the accuracy of the estimate on the direct SES factors' impacts on students' initial earnings at the same time by controlling the selectivity bias of the current employment status. The mechanism of this design is described as following.

There are two equations in the Heckman two-stage model: an earnings equation, and a selection equation. The selection equation is a "participation equation" describing the individual's propensity to work (be employed). The endogenous variables, e.g., the latent SES variables, in the selection equation are independent of the relationships, which determine the wage equation equilibrium, but nonetheless affect the equilibrium of the wage equation as well (Iyanaga & Kawada, 1980)[1]. The key insight of Heckman's approach is that observations are missing because of conscious choices (self-selection) by economic agent (e.g., the decision to work or to pursue college education). The relation between the reasons for missing observations and the nature of non-missing observations thus takes on an economic theoretical structure (The Royal Swedish Academy of Sciences, 2000)[2] and logical meaning. That is the selection bias of employment status (e.g., employed or non-employed) is yielded from the latent family socioeconomic status, which are

presented as the endogenous variables in the selection equation of the Heckit model. By calculating the significance of the Mill's Lambda of the selection equation, the statistical significance of the selectivity bias can be identified. That is, if the calculation yields significant Mill's Lambda, then it implies the existence of the statistically significant selectivity bias in the process of employment. Due to the impact of selection bias, the predictors' significance and coefficients in the earnings equation will change accordingly (comparing to the multiple linear regression equation without control of selectivity bias). Comparing the significant independent variables and their coefficients between the Heckman two-state earnings equation and the multiple regression earnings equation, the differences in between attribute to the selectivity bias causing by the latent family socioeconomic status.

To further explore the determinants of the earnings variation of the college graduates, two alternative approaches are integrated in the analytical framework. The first approach one adds gender as a predictor in the earnings equation to have gender controlled for the equation. The second approach separates the sample by gender, and applies the same Heckit model to capture the earnings characters for different gender separately.

III.3 Data

III.3.1 Data Collection and Questionnaire Design

This study uses the data set obtained from the *Survey of the Willingness on Higher Education Institution Graduates' Occupational Choice and Employment, 2003*. In order to reveal the real situation of the student-level information of college and university students, the survey was designed to have a total pool of 21600 respondents and cover samples from the whole spectrum of Chinese post-secondary education system. The survey covered different levels of post-secondary institutions in different provinces with various economic development levels. Six post-secondary institutions, e.g., two comprehensive (provincial) universities, two four-year colleges and two two-year junior colleges, were picked from each province. These six colleges and universities should represents the high end, medium and low end of post-secondary education in the selected provinces.

By original design, six provinces were chosen trichotomously from thirty one provincial level units of China, i.e., two from affluent coastal region, two from average midland region and two from underdeveloped western region. Therefore the total designed sample size is: 600x6x6=21600. In order to avoid sampling bias, the survey questionnaire was distributed randomly to students within the chosen institutions.

The stratified sampling scheme was not able to be completely implemented during the data collection. First, due to availability reason, the six provincial level units were not evenly distributed. There were three provincial level units (Beijing, Shandong Province, and Guangdong Province) from advanced costal region, two provinces (Yunnan Province and Shannxi Province) from underdeveloped western region but only 1 province (Hunan Province) from midland region. Additionally, Guangxi Province-a southwestern province voluntarily participated in the survey. Hence, the distribution of provincial unites in the final data set is: three affluent provinces, three underdeveloped provinces and one average province.

Another problem of the survey data is that the survey questionnaire is designed by the survey center, which is Institute of Economics of Education, Peking University (IEE-PKU), yet IEE-PKU did not actually print and send out the questionnaire. They sent the electronic version of questionnaire to the selected institutions and it was the institutions' job to print and distribute the questionnaire to their students. Whether the institutions exactly follow the instruction to distribute the questionnaire randomly among their senior graduates, it is not guaranteed. Yet, the descriptive statistics table of the survey data did not show obvious bias and extreme outliers, which suggests that there might not be problematic data distortion.

The survey questionnaire was given out on May 20th, 2003 and the data was reported by June 20th, 2003. The total number of returned questionnaires is 18722

(N=18722), which includes 507 voluntary samples from Guangxi Province. Therefore, filtering out the voluntary samples of 507, the response rate is 84.329% [(18722-507) / 21600].

The sample's male to female ratio is 1.375, which is higher than the national figure of 1.439 in year 2002 (MOE, 2004). To be noted that, the discrepancy between the sample figure and the national figure partially attributes to the missing value of the gender (4.68% of the observation has missing value of gender). There are 10.7% of the sample is minority in this data set, which is slightly higher than the Census 2000 figure of 8.41% (NBSC, 2003), and for this item, the missing observation takes only 4.2% of the sample.

The questionnaire is composed of 54 questions, most of which are multinomial choice questions, asking college graduates to self-report their information in the following areas: personal information, family background, study attainment, school information, job seeking situation and costs, first occupation compensations and information. Refer to Appendix I and II for the details of the questionnaire.

Though the survey data is the most current and comprehensive data available for this study, due to some design and execution problems, sampling bias could yield distorted results. First, the sampling scheme of this survey is a combination of stratified sampling and judgment sampling, which applies stratified sampling's idea

to take samples from every stratum and level of the higher education, however, identification of the relevant stratums and their actual representation in the population is not carefully quantified. The ratio of sample from each stratum (e.g., 2-2-2 provincial level sampling ratio and 2-2-2 institution level sampling ratio) was designated by arbitrary judgment of the survey designer. Second, both stratified sampling scheme and judgment sampling scheme was not implemented throughout in execution, i.e., both 2-2-2 provincial level units distribution, and 2-2-2 institution level unites distribution were not precisely carried out as designed. Therefore, the survey data might not be representative of the national population.

Sample selection bias may arise in practice for two reasons. First, there may be self selection by the individuals or data units being investigated. Second, sample selection decisions by analysts or data processors operate in much the same fashion as self selection (Heckman, 1979). In this study, both reasons of selectivity bias are designed to be reduced, by choosing various types of college and universities from three sets of provinces and, by random selection of students.

It is technically and financially hard to collect the earnings estimates from those students who didn't get job offers at that time. The survey questionnaire does not have instrumental variable to correct this truncation bias and the survey data only presents the characteristics of the college graduates.

Yet, another kind of selectivity bias is hard to eliminate by the survey

designer, which is employment selectivity. Since the data was collected by self-reporting survey among the college graduates in senior year, only those students who did enter the labor market and got job offer reported their starting wage. Though, those who didn't get job offer were in the sample, their earnings e.g., the dependent variable was missing, or say, unobservable. To address this missing data problem, this study adopts Heckman two-stage method to estimate the omitted variables' impact on the OLS model and correct the missing data error of the survey sample. The detailed earnings and selection equations will be provided in the next chapter.

III.3.2 Data Coding

Survey data were coded right after the questionnaires were retrieved. According to the properties of the questions data were coded into three kinds of variables: binary variable, ordinal variable and scale variable. The data involve a lot of students' information such as family socioeconomic status and education attainment are assessed by the questions (refer to Appendix I &II). The answers to these questions are coded into different variables in Table 4.1(in Chapter IV)

It should be noted that all the data were self-reported. There may be some error on data self-reporting, which is hard to judge. The coding of the parents'

occupation type and level is a tricky one. Referring to question number 12 in Appendix I &II, the occupation type is coded into 10 categories in a descending order for the following types: government official (equivalent to county director and up), manager, government employee, professionals, self-employed, third industry employee, business owner, industrial worker, retired \ unemployed \ semi-unemployed and, agricultural worker. For the last category—retired \ unemployed \ semi-unemployed, it involves three kinds of occupation type in fact, and they are not equal. But for the convenience of the calculation, they are coded as one category. The occupation level drops from the maximum of 10 for government official to 1 for retired \ unemployed \ semi-unemployed. The coding order is adopted from the latest social classification study done by the Academy of Social Sciences, China (Lu, 2004).

In addition, some information is not able or proper to use direct question to obtain. In such cases, indirect questions are employed to get the information. For instance, questions about "College English Test (CET) passage", "type of matriculation" and "merit-based scholarship" provide good information of students' higher education experience. CET passage is generally treated as a signal of the higher education attainment in China's current high skill labor market. For college students, Band 4 and Band 6 are the major measures. Students who pass the CET test are regarded as good in memory, language ability and communication ability. A

matriculation under national plan is considered superior to an out-of-plan matriculation, and normally means better academic standing and preparation then the latter one. And needless to say, merit-based scholarship is awarded only to students with distinguished academic performance. All of these variables reflect the position of a specific student's college experience.

III.4 Research Methodology

The key research questions are addressed by four sets of analyses. The first set of analysis is a series of statistical estimations of earnings equations using multiple linear regression. This "pilot" analysis identifies significant factors of initial earnings of the college graduates with traditional statistical technique and without correction for selection bias causing by family socioeconomic status. The second set of analysis employs Heckman two-stage method (or Heckit Model). The purpose of this set of analysis is to utilize the selection equation (equation 2) of the Heckit model to correct the selectivity bias caused by family socioeconomic status, then compare the results from the earnings equation (equation 1) of the Heckit model to the previous multiple regression result. The differences of the earnings equations between the pilot analysis and Heckit model reveals the differences due to selectivity

bias. The third set of analysis is the alternative estimation of the earnings model with Heckman two-stage method, which focuses on gender as a major alternative explanation to the earnings gaps among college graduates. The fourth set of analysis employs a probit approach to explain the family socioeconomic status's impact on initial earnings through employment.

III.4.1 Pilot Analysis Using Multiple Regression

Multiple regression estimates the coefficients of the linear equation, involving one or more independent variables, which best predict the value of the dependent variable. The linear regression model assumes that there is a linear, or "straight line," relationship between the dependent variable and each predictor. This relationship is described in the following formula.

$$y_i = b_0 + b_1 x_{i1} + b_j x_{ij} + \ldots + b_p x_{ip} + e_i$$

Where

y_i is the value of the i^{th} case of the dependent scale variable

p is the number of predictors

b_j is value of the j^{th} coefficient, j=0,…,p

x_{ij} is the value of the i^{th} case of the j^{th} predictor

e_i is the error in the observed value for the i^{th} case

Note that b_0 is the intercept, the model-predicted value of the dependent variable when the value of every predictor is equal to 0.

Specific to this study, the initial annual earnings in Yuan per year (EXPANNER) is the dependent variable *Y*. For independent variables: overall class ranking, College English Test (CET) score, scholarship and popular major are [*i*] vector of schooling variables and parents' education level, parents' occupation level, job seeking cost, miscellaneous expenses, family financial assistance, student loan, origin type and working experience are [*j*] vector of control variables.

Detailed description of the earnings equation will be present in the next chapter.

III.4.2 Heckman Two-stage Method Equation

The application of Heckman two-stage method or Heckit model on economics of education is quite common. The difficulty of determine the returns to higher education investment are to characterize the counterfactual situation, i.e., to answer the question: what would have happened if the individual had not participated in the program? Since it is impossible to observe an individual as both employed and unemployed, it is necessary to use information on labor market's outcomes for this purpose. Given that the allocation of individuals to higher

education is seldom purely random, the group of college graduates becomes a selected sample with observed and unobserved characteristics that may differ from those of the overall population. The main ingredients of Heckit method are twofold: (1) a model of production, and (2) a model of selection.

According to the model for this study, college graduates' initial earnings (wage) Y could be correlated with X, a vector of exogenous independent variables including individual characteristics, family socioeconomic status, school characteristics, etc. d is a binary dummy variable, d=1 if the student has an job offer and d=0 if the student does not find a job. Hence,

(1) $Y = C + X\beta + d\alpha + \epsilon$

Let d^* be the latent selection variable which, when $d^* >= 0$, d=1 when $d^* < 0$, d=0. Selectivity into the labor market is believed to be related with Z, a vector of family and individual attributes and characteristics. Then:

(2) $d^* = Z\delta + v$

ϵ is assumed to be independent of X and Z. $E(\epsilon)=0$, $E(v)=0$ and $\text{var}(v)=1$. The joint distribution of ϵ and v conditional on X and Z is assumed bivariate normal, N (0, 0, $\sigma 2\ \epsilon$, 1, ρ), where ρ is the correlation between ϵ and v. If the disturbances in a regression model are correlated with the disturbances in the selection equation, conventional estimation techniques will not provide consistent estimates of the parameters (Lee, 1982)[3]. E ($Y_e - Y_u$) without adjusting for sample selection gives E

$(Y_e \mid d^* >= 0) - E(Y_u \mid d^* < 0) = \alpha + [E(\epsilon \mid d = 1) - E(\epsilon \mid d = 0)]$, where $[E(\epsilon \mid d = 1) - E(\epsilon \mid d = 0)]$ captures the selection effect and α the treatment effect. Therefore, estimating α by taking the difference of observed earnings of the hired student and those students without job offer recover the real α except when selection is random.

Lee (1978)[4], Heckman (1979), Lee (1982) and other authors built a framework of switching regression models, and the equation (1) can be reconstructed as the following:

(3) $Y_e = X_e\beta_e + \epsilon_h$

(4) $Y_u = X_u\beta_u + \epsilon_u$

Where equation (3) is the employment equation and equation (4) is the unemployment equation, $\epsilon_e \sim N(0, \sigma 2_e)$ and $\epsilon_u \sim N(0, \sigma 2_u)$. Subscript e stands for employed and u for unemployed.

For those unemployed students, their wage if they had employed is not observable. Therefore, equation (3) is conditional on employment, and similarly equation (4) is conditional on unemployment.

(5) $E(Y_e \mid X_e, selection) = X_u\beta_u + E(\epsilon_e \mid d^* >= 0)$

In general, $E(\epsilon_e \mid d=1) \neq 0$ and $E(\epsilon_u \mid d=0) \neq 0$. Under the assumption of bivariate normal distribution of ϵ and ν[1], the employment and the unemployment

[1] Lee (1982) pointed out that the selectivity bias terms in the regression equation may be sensitive to the specific probability models even though there may be only slight differences in the probability models. Lee's paper addressed the general applicability of this selectivity bias correction approach so that non-normal conditions can still be managed.

regressions can be consistently estimated by following a Heckman two-stage least squares procedure by treating selectivity as a "missing variable" (refer to Heckman, 1979 for detailed derivation). The estimated earnings from the two equations allow unbiased comparison of family SES and individual characteristics between employed and unemployed students. The Mill's Lambda or Inverse Mill's Ratio (IMR) is calculated in the Heckman two-stage model. In fact the Inverse Mill's Ratio is calculated as an explanatory variable containing the nonselection hazard. The nonselection hazard is derived from a probit regression of whether the dependent variable is select or observed. Under full maximum likelihood procedure, the nonselection hazard is derived from the parameter estimates of the selection equation (Heckman, 1979).

III.4.3 Gender Analysis with Heckman Two-stage Method

So far this study is focused on the correlation between college graduates' family socioeconomic status and their initial annual earnings. The previous two sections concentrate on identifying and capturing the existence and magnitude of the family socioeconomic status' impact on earnings. However, apart from higher education experience and family socioeconomic status, there are other factors that may also influence earnings differences among the college graduates. The alternative

models are constructed to explore the impact on earnings based on gender differences.

In fact, several alternative approaches can be considered. Numerous literatures and studies provided conceptual frameworks for the alternatives. Gender, race, household composition, parental wealth and other factors, which are not included in the preceding two models, could all be considered as the alternatives to explain the earnings differences. However, this study concentrates on gender as a factor, due to several empirical and institutional rationales.

First of all, although race is an important determinant in earnings equation (Card & Krueger, 1992), it is not a sizable consideration in this study. According to the latest national census (2000), 91.59 percent of the total population is of Han nationality, which falls into the category of Mongol (or Yellow Race) in ethnography; and 8.41 percent were various national minorities, most of which are also classified as Yellow Race. There is no Negro (Black Race) and Caucasian (White Race) population reported in China's census ever. Therefore, race is not applicable to this study.

Many researchers like Becker (1966), Rumberger (1983) and Behrman & Taubman (1986) argue that birth order or sibling number is an important determinant of an individual's earnings. However in China, due to the mandatory one-child policy enacted in the late 1970s, household composition (numbers of siblings) is not

a consideration in this study too. The survey was conducted in 2003 for college seniors, who are mainly born after 1980. That is, most of the survey participants are not likely to have any siblings in their generation. This is another unique scenario of China case.

For family socioeconomic status characters, family wealth could be a very important factor in earnings prediction (Rumberger, 1983; Campbell & Henretta, 1980). However, family wealth is very hard to measure and quantify (Rumberger, 1983). Especially in the case of China, a lot of students do not know how much wealth their families possess in monetary form (Bian & Logan, 1996). Additionally, the survey questionnaires are filled out by the college graduates voluntarily, such question about family wealth and family gross income were excluded from the design because they are highly germane to personal privacy. Therefore, technically, data regarding to family wealth are not obtainable for this study. Consequently, a study of family wealth is not feasible.

Though the alternatives mentioned above are either not applicable or technically unfeasible, an alternative model based on gender is feasible and applicable. Gender as a determinant in the earnings equation is conducted as the third set of analysis. China has been promoting the equal opportunity of male and female in vast aspect of social lives even since the founding of People's Republic of China in 1949. Yet the real effects of such effort are not examined with carefully

designed quantitative methodology until recently (Wen, 2004).

The alternative model involves two sets of Heckman two-stage model analyses. The first set is similar to the second set of analysis described above; the only difference is that gender is put into the earnings equation (equation I) to estimate gender's impact on earnings. The second set divides the sample into two subgroups by gender, and then applies the same Heckit model equations in the second set of analysis to estimate the earnings determinants for different gender groups separately. The mechanism of this two-folded designed alternative model is that: when the variable of gender is added into the earnings equation, it will control the impact of gender to the earnings and change the significance and magnitudes of the determinants in the earnings equations. Due to the collinearity the effect of the other determinants on earnings will either enlarges or shrinks accordingly. When the sample is divided into two subgroups by gender, the effect of gender is eliminated entirely. Within each sub-sample, then the same Heckit model is applied to each of the sub-sample to estimate the earnings determinants. From this alternative model, the independent variables take effect on the earnings exclusively from the impact of gender and the results can depict a better picture of the influence of the predictors on earnings.

III.4.4 Explain the Mechanism of SES's Impact on Earnings with a Probit Approach

The probit approach is employed to reveal the mechanism between various factors, especially the family socioeconomic status factors' impact on college graduates' initial earnings. The hypothesis is: various factors (including SES factors) influence the propensity of a college student to get a job offer after graduation. Due to the different financial pressure and aspiration of getting a high pay job or continuing graduate education, family SES factors take effect in the process of employment decision making and yield difference in first job offer salary, e.g., initial earnings.

Technically, probit regression is an approach to handling categorical dependent variables. A typical use of probit approach is to analyze dichotomous dependent variable, specific to this study, the dependent variable is current employment status (CES), which has a pair of dichotomous values 1 and 0 representing employed and non-employed. Probit regression focuses on a transformation of the probability that Y, the dependent variable, equals 1. The function used in probit is the inverse of the standard normal cumulative distribution function. Probit regression assumes the categorical dependent variable reflects an underlying quantitative variable and it uses the cumulative normal distribution.

The equation of probit model is defined as:

$Pr(y=1|x) = \Phi\ (xb)$

where Φ is the standard cumulative normal probability distribution and xb is called the probit score or index.

Since xb has a normal distribution, interpreting probit coefficients requires thinking in the Z (normal quantile) metric. The probit coefficient is how much difference a unit change in the independent makes in terms of the cumulative normal probability of the dependent variable. This means the probit coefficient measures the effect of the independent on the Z scores of the dependent. Note that the probability of the dependent is not a linear function of Z, but rather is a cumulative normal function of Z. This means that the effect of a unit change in the independent on the probability of the dependent depends on the level of the independents. Therefore to assess the effect of probit coefficients it is necessary to choose some level of the independents as a reference point and in particular the standard reference point is when all independents are at their sample means. In practical terms, the cumulative standard normal probit curve displays an S-shaped curve, which rises from 0 to 1.

CHAPTER IV

ANALYSIS OF DATA

Six types of analysis are conducted in this study: (1) descriptive statistic analysis—to explain data cleaning process and present general statistics of the sample; (2) mean / median comparison tests—to divide sample into subgroups, compare subgroups' means and medians, use both linear and non-parametric test (including Kruskal Wallis test and median test) to compare the differences in sub-sample means and medians; (3) multiple regression analysis to estimate the determinants of initial annual earnings of college graduates; (4) Heckman two-stage consistent estimate method to estimate earnings equations with self-selection bias corrected ; (5) estimate the initial annual earnings equations using alternative model specifications and sub-samples—to explore alternative explanations to the variations of college graduates' initial earnings other than family SES, and (6) estimate employment status equation using a probit model.

IV.1 Data Cleaning and Basic Characteristics of the Sample

All variables in the sample are sorted and checked for unreasonable extreme values. Since all cases with larger values appear on the top of the data set, it is easier to check for outliers. It should be noted that because some variables can be "0" or even missing values, an ascending order is not preferred in the extreme outliers censoring in this study. After sorting, some records are considered as obvious unreasonable outliers and cleared. For instance, there are 5 cases reported a 20-digit College Entrance Examination (CEE) score, which are absolutely unreasonable; 1 case has a reported family financial assistance of 10 million Yuan; 3 cases reported the student loan amount more than 130,000 Yuan; 4 cases reported the food expenses more than 100,000 Yuan. These cases are treated as unreasonable outliers and deleted from the data set. Also, there are several cases that have very high reported scholarships. These cases are also quite problematic. In China, only a few elite universities offer large amount of scholarships to students, and they are exclusively on merit base. Therefore, these high-volume scholarship cases' CEE score and overall class ranking are examined. Those cases with high scholarship but low CEE score and class ranking are considered as extreme outliers, and deleted.

In addition, this study only focuses on college graduates; therefore, all cases with other degree levels (e.g., graduate level and associate level) are filtered out from the data set. After the trimming process, 10146 records in total are considered valid, which filters out 8576 cases from the original 18722 records. Table 4.2 shows the basic statistics of the trimmed data set. Variable name, number of valid record, mean, median, minimum and maximum are reported. It should be noted that some of the variables are categorical or ordinal, the median is computed instead of the mean. Table 4.1 is a code book of the variables, which lists the names of the variables and their labels.

Table 4.1 Variables Names and Labels

Variable Name	Label (with unit and explanation)
ID	ID number of the record
MAJOR	Code of the enrolled major/program
ORGNTYP	Origin type (from what type of resident locality)
CEE	College entrance examination (CEE) score
CEETP	CEE score type (raw score or standardized score)
FATHOCCP	Father's occupation type
MOTHOCCP	Mother's occupation type
FHSL	Father's highest schooling level
MHSL	Mother's highest schooling level
RANK	Overall class ranking (self-reported)
WKEXPCBC	working experience before college
WKEXPC	working experience
SCHSHIPA	Amount of scholarship (¥, in total)
FAMILYAS	Family financial assistance amount (¥, in total)
LOANAMNT	Amount of student loan (¥, in total)
FDEXPNS	Food expenses (¥, in total)
MISCLEXP	Miscellaneous expenses (¥, in total)
EXPANNER	Expected annual earnings by the end of survey (¥/year)

RSVWAGE	Reservation wage(￥/year)
FOLLOWUP	Follow-up survey participation
FATHSCHL	Father's schooling years
MOTHSCHL	Mother's schooling years
ORIGNPRV	Origin province of the student
MATRITYP	Matriculation Type (Under national plan vs. out-of-plan)
DEGREE	Degree enrolled (associate/bachelor/graduate)
OWNSHP	Ownership of the institution (Public vs. non-public)
SEX	Gender
CES	Current employment status (employed/non-employed)
SCHLP	Scholarship awarded or not
LOAN	Having student loan or not
CES2	Current employment status in follow-up web-survey
POPMAJ	Popular major or not
FATHEDLV	Father's Education level (above or below median level)
MOTHEDLV	Mother's Education level (above or below median level)
FATHOCCL	Father's Occupation Level (above or below median level)
MOTHOCCL	Mother's Occupation Level (above or below median level)
Ethny	Ethnicity (Han or Non-Han)
ORGNLV	Origin Type Recode (above or below median level)

Table 4.2 Basic Statistics of Data Set

	Mean			Median			Min	Max	N
	Female	All	Male	Female	All	Male			
ExpAnnErn	21133.56	21103.12	21144.39	20000.00	20000.00	20000.00	0	1000000	4012
FthOccTy	N/A	N/A	N/A	4.00	3.00	2.00	1	10	9100
MothOccTy	N/A	N/A	N/A	2.00	2.00	2.00	1	10	9802
FHSL	N/A	N/A	N/A	4.00	4.00	4.00	1	7	9745
MHSL	N/A	N/A	N/A	5.00	5.00	5.00	1	7	9611
FathSchlYr	11.60	11.12	10.83	12.00	12.00	12.00	2	19	9745
MothSchlYr	9.92	9.21	8.77	9.00	9.00	9.00	2	19	9611
CET Pass	1.17	1.07	1.00	1.00	1.00	1.00	0	2	9787
PopMaj	N/A	N/A	N/A	1.00	1.00	1.00	0	1	9820
OrgnTy	N/A	N/A	N/A	2.00	2.00	2.00	1	4	9870
WkExpc	N/A	N/A	N/A	.00	.00	.00	0	2	9529
Rank	N/A	N/A	N/A	2.00	2.00	2.00	1	4	9717

SchshipAmn	1466.85	1295.94	1189.22	800.00	500.00	500.00	0	40000	6777
LoanAmnt	953.46	1472.39	1762.72	.00	.00	.00	0	50000	5693
JbSkCat	942.01	920.14	909.36	500.00	500.00	470.00	10	10000	8489
MatrTyp	N/A	N/A	N/A	1.00	1.00	1.00	0	1	10138
OwnShp	N/A	N/A	N/A	1.00	1.00	1.00	0	1	10125
FamilyAssist	28357.70	27046.47	26435.09	30000.00	30000.00	30000.00	0	300000	6412
MiscExpns	7352.90	6761.22	6433.32	5000.00	4400.00	4000.00	0	150000	6424
CES	N/A	N/A	N/A	.00	1.00	1.00	0	1	8692
Ethnicity	N/A	N/A	N/A	1.00	1.00	1.00	0	1	10137

The raw data set is also recoded for the purpose of analysis. FathOccpTy, MothOccpTy, FHSL, MHSL and OrgnTyp are recoded into boolean variables FathOccLv, MothOccLv, FathEdLv, MothEdLv and OrgnLv.

Table 4.3 Recoding of Variables

Original variable	Variable label	Data type	Median	Recoded variable	Data type	Recoding formula
FHSL	Father's highest schooling level	Ordinal	4	FathSchlYr	scale	7=0 6=6 5=9 4=12 3=15 2=18 1=19
MHSL	Mother's highest schooling level	Ordinal	5	MothSchlYr	scale	
FHSL	Father's Education level	Ordinal	4	FathEdLv	binary	1~3=1 4~7=0
MHSL	Mother's Education level	Ordinal	5	MothEdLv	binary	1~4=1 5~7=0
FathOccpTyp	FathOccTy	Ordinal	8	FathOccLv	binary	1~7=1 8~10=0
MothOccpTyp	Mother's occupation type	Ordinal	9	MothOccLv	binary	1~8=1 9~10=0
Major	Major	categorical	n/a	PopMaj	binary	10,2,3&11=1 other=0
OrgnTyp	Origin type	Ordinal	2	OrgnLv	binary	1=1,2&3=0

The variables of FathOccTy and mother's occupation type are coded in a

descending order. According to the recent social stratum classification report done by China Academy of Social Science (Lu, 2004), the occupation types are coded as: 1-Government official (equivalent to county director and up), 2-Manager, 3-Government employee, 4-professionals, 5-Self-employed, 6-Third industry employee, 7-Business owner, 8-Industrial worker, 9-Retired \ Unemployed \ Semi-unemployed and, 10-Agricultural worker. The levels of occupation type are descending accordingly.

IV.2 Mean/Median Comparisons for Different Population Sub-groups

IV.2.1 Simple Tabulation and Linear Test

The first part of analysis focus on structural heterogeneity of the samples from different population groups and compare the means or medians of the key characteristics. These 5 key characteristics are: father's occupation type (FATHOCCP), mother occupation type (MOTHOCCP), father's highest schooling level (FHSL), mother's highest schooling level (MHSL) and origin type (ORGNTYP). In this data set, the first four of these variables are most common

used indicators in family socioeconomic status studies. The last one, origin type, is also an important family SES background indicator, especially in China. The variable of origin type has 3 ordinal values 1, 2, and 3, which stand for village, county and city separately (please refer to the Appendix I and II). The permanent residency registration system, so called *Hu Kou* is still functioning in China, which prohibits free migration from rural areas to the cities. It gives useful information about a student's family socioeconomic status. Generally speaking, the family socioeconomic status rises from village, to county and, to city.

First, let's take a look at the two scale variables of SES characters-father's schooling years and mother's schooling years. The means of sub-samples are computed.

Table 4.4 Means of Parents' Schooling Years

		Mean schooling years	
		Father	Mother
Mother's Occupation Level	< = median	9.84	7.64
	> median	13.16	11.76
Father's Occupation Level	< = median	9.18	7.47
	> median	13.11	11.01
Origin Level	< = median	10.27	8.16
	> median	13.04	11.59
CET Passage	Fail or no score	10.71	8.74
	Band 4	11.02	9.12
	Band 6	11.55	9.65
working experience	No	11.11	9.14
	Yes	11.06	9.27
Popular major	No	10.64	8.90
	Yes	11.18	9.29

Matriculation Type	Out-of-plan	12.01	9.87
	National plan	11.10	9.19
Ownership of the institution	Non-public	11.52	9.72
	Public	11.12	9.20

Table 4.4 shows that in terms of father and mother's schooling years, there are no differences between two subgroups if the sample is divided by the popularity of major and current employment status. Yet, if the sample is divided by matriculation type and ownership of the institution, the mean parental schooling years are different between groups. However, the intuitive differences are yet to be tested by linear regression to see if the differences are statistically significant. The following tables show the mean comparison t-test results of these subgroups.

Table 4.5 Mean Comparison T-test by Matriculation Type

	Matriculation Type	N	Mean	Std. Deviation	Std. Error Mean
Father's schooling years	0	256	12.01	3.539	.221
	1	9480	11.10	3.605	.037
Mother's schooling years	0	244	9.87	3.879	.236
	1	9360	9.19	3.663	.040

		t-test for Equality of Means						
		t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
							Lower	Upper
Father's schooling years	Equal variances assumed	4.000	9734	.000**	.913	.228	.466	1.360
	Equal variances not assumed	4.071	269.487	.000**	.913	.224	.471	1.354
Mother's schooling years	Equal variances assumed	2.729	9602	.006**	.685	.251	.192	1.178
	Equal variances not assumed	2.866	257.309	.004**	.685	.239	.214	1.155

Table 4.5 shows parental schooling years are significant different between

groups by matriculation type.

Table 4.6 Mean Comparison T-test by Ownership of Institution

	Ownership of Inst.	N	Mean	Std. Deviation	Std. Error Mean
Father's schooling years	0	136	11.62	3.885	.333
	1	9589	11.12	3.602	.037
Mother's schooling years	0	132	9.72	3.496	.304
	1	9480	9.20	3.684	.040

		t-test for Equality of Means						
		t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
							Lower	Upper
Father's schooling years	Equal variances assumed	1.288	9723	.198	.401	.311	-.209	1.012
	Equal variances not assumed	1.197	138.312	.234	.401	.335	-.262	1.064
Mother's schooling years	Equal variances assumed	1.521	9590	.128	.517	.340	-.149	1.184
	Equal variances not assumed	1.685	135.547	.094	.517	.307	-.090	1.124

Table 4.6 shows that there are no significant differences in terms of parents' schooling years between the groups of public and non-public institutions.

Table 4.7 Mean Comparison T-test by Popularity of Major

	Popular Major	N	Mean	Std. Deviation	Std. Error Mean
Father's schooling years	0	2883	10.84	3.673	.069
	1	6604	11.16	3.533	.043
Mother's schooling years	0	2851	8.90	3.968	.074
	1	6499	9.29	3.811	.047

		t-test for Equality of Means						
		t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
							Lower	Upper
Father's schooling	Equal variances assumed	-4.222	9485	.000**	-.338	.080	-.495	-.181

years	Equal variances not assumed	-4.156	5252.59	.000**	-.338	.081	-.497	-.179
Mother's schooling years	Equal variances assumed	-4.436	9348	.000**	-.385	.087	-.555	-.215
	Equal variances not assumed	-4.387	5246.21	.000**	-.385	.088	-.557	-.212

Table 4.7 shows that the differences of parents' schooling years are statistically significant between students enrolled in popular majors and those who enrolled in non-popular majors.

Table 4.8 Mean Comparison T-test by Current Employment Status (CES)

	C.E.S.	N	Mean	Std. Deviation	Std. Error Mean
Father's schooling years	0	4434	11.35	3.627	.054
	1	5213	10.92	3.574	.049
Mother's schooling years	0	4377	9.41	3.927	.059
	1	5139	9.02	3.627	.053

		t-test for Equality of Means						
		t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
							Lower	Upper
Father's schooling years	Equal variances assumed	5.950	9645	.000**	.437	.074	.293	.581
	Equal variances not assumed	5.943	9352.95	.000**	.437	.074	.293	.582
Mother's schooling years	Equal variances assumed	4.960	9514	.000**	.395	.080	.239	.551
	Equal variances not assumed	4.950	9194.53	.000**	.395	.080	.239	.552

Table 4.8 shows that the differences of parents' schooling years are statistically significant between employed and yet-to-be-employed graduates.

Also, the mean earnings of students from different SES background are quite different.

Table 4.9 Mean Expected Annual Earnings by Different Factors

		Expected annual earning
		Mean (Yuan / year)
Father's Occupation Level	0 (< = median level)	19362
	1 (> median level)	23206
Mother's Occupation Level	0 (< = median level)	19699
	1 (> median level)	24019
Father's Education Level	0 (< = median level)	20104
	1 (> median level)	24161
Mother's Education Level	0 (< = median level)	19724
	1 (> median level)	23714
Origin Level	0 (< = median level)	19811
	1 (> median level)	24868
CET Passage	0 (Fail or no score)	18603
	1 (Band 4)	21783
	2 (Band 6)	24460
working experience	No	21771
	Yes	19334
Popular major	No	19587
	Yes	21894
Scholarship	No	20348
	Yes	21795
Student loan	No	21387
	Yes	19984
Matriculation Type	Out-of-plan	17004
	Under national plan	21155
Ownership of the Institution	Non-public	13643
	Public	21137
Gender	Female	21134
	Male	21144

Table 4.9 shows obvious gaps of initial annual earnings by different characters. Browsing through the table, almost all characters except gender seems to have major impact on earnings. The earnings seems quite sensitive to family

socioeconomic status factors like parents' education level, occupation level and origin type level, yet pretty egalitarian in terms of gender (only 10-Yuan difference). However, this tabulation presents only means of each sub-group, without applying any statistical test and bias control, which means the first impression could be wrong.

IV.2.2 Nonparametric Tests

Apart from a few scale variables like father's schooling years and mother's schooling years, quite a few other socioeconomic status characteristics are categorical variables. Mean comparison t-test is not suitable for these variables. The nonparametric tests for multiple independent samples are useful for determining whether these categorical variables differ between two or more groups. This is especially true when the assumptions of linear regression are not met. Specific to this study, several test variables are ordinal; the mean is not a valid estimate because the distances between the values are arbitrary. Even if the mean is valid, the distribution of the test variable might not be normal. Two types of nonparametric tests are conducted here: the median test[2] and the Kruskal-Wallis

[2] The median test is designed to test the null hypothesis that groups have the same median. Because the test makes no assumptions about the data other than that the median is a valid measure of center, it can be used in

test.[3]

Obviously, the distribution of the cases here is not likely to be random by these grouping variables. Whether a student goes to a public university under the national plan, or whether a student is admitted by a popular major is not random, or say, independent to other factors. The null hypothesis for this test of tests is that: student's matriculation, college choice, major choice and current employment status have no significant difference by their family socioeconomic status.

IV.2.2.1 Nonparametric test by different matriculation types: the

a variety of situations. It is especially appealing when the test variable has different distributions by group. One weakness of the test is that it is not designed to take advantage of distance from the median. When the assumptions behind the standard ANOVA are invalid or suspect, the nonparametric procedures are good alternatives to test for the significance of the difference between multiple groups. The null hypothesis for the median test is that this particular median value is a good approximation of center for each group. To test this hypothesis, each group is divided into two subgroups: those whose value is equal to or below the median, and those whose values are above it. The result is a two-way frequency table with two rows and g columns, where g is the number of categories in the grouping variable. For incidence in this study, g=2, e.g., the categories in the grouping variable equal to 2. That is, matriculation type (under national plan vs. without national plan), type of institution (public vs. non-public), type of major (popular major vs. non-popular major) and current employment status (employed vs. unemployed). In fact, the median test is a chi-square test of independence between groups and the proportion of cases above and below the median. The chi-square value is obtained in the usual fashion for two-way tables. For each cell, the distance between the observed and expected counts is squared, and then divided by the expected value. Finally, these quantities are summed across all cells. The asymptotic significance shows how often a chi-square value at least as large as Chi-Square value in similar repeated samples is expected. The median test is designed to test the null hypothesis that groups have the same median. Because the test makes no assumptions about the data other than that the median is a valid measure of center, it can be used in a variety of situations.

[3] The Kruskal-Wallis test uses ranks of the original values and not the values themselves. That's appropriate in ordinal variable case, because the scale used by the test is ordinal. If you would like to take advantage of these distances and can assume that the groups have similar distributions on your test variable, then you should consider using the Kruskal-Wallis test. The Kruskal-Wallis test is a popular nonparametric alternative to the standard one-way analysis of variance. It is appropriate when your test variable is ordinal or its distribution does not meet the assumptions of standard ANOVA. The only assumptions made by the test are that the test variable is at least ordinal and that its distribution is similar in all groups. First, each case is ranked without regard to group membership. Cases tied on a particular value receive the average rank for that value. After ranking the cases, the ranks are summed within groups. The Kruskal-Wallis statistic measures how much the group ranks differ from the average rank of all groups. The chi-square value is obtained by squaring each group's distance from the average of all ranks, weighting by its sample size, summing across groups, and multiplying by a constant. The degrees of freedom for the chi-square statistic are equal to the number of groups minus one. The asymptotic significance estimates the probability of obtaining a chi-square statistic greater than or equal to the one displayed, if there truly are no differences between the group ranks. A chi-square value with the same value of the degrees of freedom should occur only about the chances of the value of reported asymptotic significance.

sample is divided into two groups according to "matriculation type" of the students, e.g., matriculation under national college enrollment plan or not, and examine if there is any different characteristics of the socioeconomic status between two groups.

Table: 4.10 Kruskal Wallis Test Statistics by Matriculation Type (a)

	Origin type	FathOccTy	Mother's occupation type	Father's highest schooling level	Mother's highest schooling level
Chi-Square	20.303	32.803	50.268	15.583	7.442
Df	1	1	1	1	1
Asymp. Sig.	.000	.000	.000	.000	.006

a Grouping Variable: Matriculation Type

Table: 4.11 Median Test Statistics by Matriculation Type (a)

		Origin type	FathOccTy	Mother's occupation type	Father's highest schooling level	Mother's highest schooling level
N		9861	9706	9595	9736	9604
Median		2.00	8.00	9.00	4.00	5.00
Chi-Square		11.905	44.170	41.524	5.336	10.135
df		1	1	1	1	1
Asymp. Sig.		.001	.000	.000	.021	.001
Yates' Continuity Correction	Chi-Square	11.467	43.325	40.670	5.046	9.706
	df	1	1	1	1	1
	Asymp. Sig.	.001	.000	.000	.025	.002

A Grouping Variable: Matriculation Type

Table 4.10 and 4.11 show that all family socioeconomic status characteristics are significantly different between the two groups. The Asymp.Sig. values in both tables are significant. It also indicates that the medians of origin type, father's occupation type, mother's occupation type, father's highest

schooling level and mother's highest schooling level for the whole population are: county, industrial worker, retired \ unemployed \ semi-unemployed, high school or equivalent and junior high separately.

It should be noted that, the asymptotic significance values in the Kruskal Wallis test table (Table 4.10) estimate the probability of obtaining a chi-square statistic greater than or equal to the one displayed in the table, if there truly are no differences between the group ranks. Chi-square of 20.303, 32.603, 50.268, 15.583 and 7.442 with 1 degree of freedom is reported to have very low chance to occur in Table 4.10 (i.e., Asymp. Sig. < 0.05). Hence the null hypothesis of Kruskal Wallis test is rejected, in other words, the SES differences between two groups are significant.

For the median test, the Asymp.Sig. values indicate there are almost no chance (Asymp.Sig. = 0) that a chi-square value at least as large as reported values (e.g., 11.905, 44.170, 41.524, 5.045 and 10.135) in similar repeated samples, if there really is no relationship between the median and group membership. The median test result shows that this probability is very low: for instance, "Mother's highest schooling" has only 1 out of 1000 chances. Therefore, the null hypothesis is rejected; there exist significant socioeconomic status differences between two groups.

IV.2.2.2 Nonparametric test by different type of institutions:

According to the "ownership of institution", e.g., public or non-public, the sample is divided into two groups and examined if the SES characteristics are different.

Table 4.12 Kruskal Wallis Test Test Statistics by Ownership of Institution (a)

	Origin type	FathOccTy	Mother's occupation type	Father's highest schooling level	Mother's highest schooling level
Chi-Square	9.532	10.524	13.319	1.756	2.746
df	1	1	1	1	1
Asymp. Sig.	.002	.001	.000	.185	.097

a Grouping Variable: Ownership of the institution

Table 4.13 Median Test Statistics by Ownership of Institution (a)

		Origin type	FathOccTy	Mother's occupation type	Father's highest schooling level	Mother's highest schooling level
N		9851	9695	9583	9725	9592
Median		2.00	5.00	9.00	4.00	5.00
Chi-Square		2.719	8.627	18.746	.232	4.555
df		1	1	1	1	1
Asymp. Sig.		.099	.003	.000	.630	.033
Yates' Continuity Correction	Chi-Square	2.439	8.122	17.969	.155	4.170
	df	1	1	1	1	1
	Asymp. Sig.	.118	.004	.000	.694	.041

a Grouping Variable: Ownership of the institution

Table 4.12 shows that both father's and mother's highest schooling level are not significantly different between two groups. In addition, median test concurs the result of Kruskal Wallis test on parents' highest schooling level and even identify the insignificance of origin type. Table 4.13 shows that father's occupation type, mother's occupation type and mother's highest schooling level are significantly different at 5% level between the two groups; however the origin type

and father's highest schooling level are not significant, which indicates the difference of these two variables between two groups is not statistically significant, or say, there are no difference between two groups in term of father's highest schooling level and origin type.

IV.2.2.3 Nonparametric test by different major: The sample is split into two groups to test if students in different majors, e.g., popular major or non-popular major have different SES characteristics.

Table 4.14 Kruskal Wallis Test Statistics by Popularity of Major (a)

	Origin type	FathOccTy	Mother's occupation type	Father's highest schooling level	Mother's highest schooling level
Chi-Square	9.089	11.303	9.789	15.773	17.849
df	1	1	1	1	1
Asymp. Sig.	.003	.001	.002	.000	.000

a Grouping Variable: Popular major

Table 4.15 Median Test Statistics by Popularity of Major (a)

		Origin type	FathOccTy	Mother's occupation type	Father's highest schooling level	Mother's highest schooling level
N		9598	9440	9344	9467	9350
Median		2.00	8.00	9.00	4.00	5.00
Chi-Square		8.793	12.045	12.793	12.356	21.677
df		1	1	1	1	1
Asymp. Sig.		.003	.001	.000	.000	.000
Yates' Continuity Correction	Chi-Square	8.661	11.889	12.629	12.197	21.456
	df	1	1	1	1	1
	Asymp. Sig.	.003	.001	.000	.000	.000

a Grouping Variable: Popular major

Table 4.14 and 4.15 show the same pattern that all medians of the five

socioeconomic status indexes are identical with the preceding tests; all family socioeconomic status characters are significantly different at 5% level between the two groups.

IV.2.2.4 Nonparametric test by current employment status: The sample is split into two groups by whether the respondent gets a job offer by the end of survey to see if there is any difference in SES background between two groups.

Table 4.16 Kruskal Wallis Test Statistics by Current Employment Status (a)

	Origin type	FathOccTy	Mother's occupation type	Father's highest schooling level	Mother's highest schooling level
Chi-Square	43.403	35.993	51.745	39.856	25.461
df	1	1	1	1	1
Asymp. Sig.	.000	.000	.000	.000	.000

a Grouping Variable: Current employment status

Table 4.17 Median Test Statistics by Current Employment Status (a)

		Origin type	FathOccTy	Mother's occupation type	Father's highest schooling level	Mother's highest schooling level
N		9730	9613	9506	9647	9516
Median		2.00	5.00	9.00	4.00	5.00
Chi-Square		26.610	28.682	67.383	16.921	19.701
df		1	1	1	1	1
Asymp. Sig.		.000	.000	.000	.000	.000
Yates' Continuity Correction	Chi-Square	26.598	28.472	67.038	16.751	19.509
	df	1	1	1	1	1
	Asymp. Sig.	.000	.000	.000	.000	.000

a Grouping Variable: Current employment status

Table 4.16 and 4.17 again show the similar propensity as preceding tests

that all medians of the socioeconomic status characteristics are identical with the preceding tests; all family socioeconomic status characters are significantly different at 5% level between the two groups.

Obviously, the nonparametric analysis results above demonstrate the heterogeneity of different groups in family socioeconomic status by matriculation type, ownership of the institution, major and current employment status. The median test and Kruskal-Wallis test also confirm the result of preceding mean comparison t-test.

The preceding tests also portrait the typical feature of the family socioeconomic status of the majority student body, which is: coming from county level origin, with a high-school-level educated father working as an industrial worker and a junior-high-school educated mother retired or unemployed

IV.3 Determinants of Initial Earnings

IV.3.1 Linear Regression Model

Without any predefined or expected outcome, a pilot regression is conducted first to reveal the correlations among predictors and dependent variable. The independent variables of the pilot regression, e.g., the 14 predictors fall into

three categories: higher education experiences, family socioeconomic status and labor market preferences. These 14 variables are: CET Passage, Class ranking and amount of scholarship are treated as higher education experiences predictors; miscellaneous expenses, amount of student loan, Family financial assistance amount, origin level, father's occupation level, father's education level, mother's education level and mother's occupation level are treated SES related predictors; popular major and working experience are treated as labor market preferences predictors. These predictors are chosen due to the intuitive correlation to the dependent variable. This preliminary regression turned out to fit the sample well.

Table 4.18 Multiple Regression Model Fit (b)

	Sum of Squares	df	F	Sig.
Regression	2.09E+10	14	14.948	.000(a)
Residual	1.34E+11	1343		
Total	1.55E+11	1357		

a Predictors: (Constant), Mother's Education Level, working experience, Amount of scholarship, Miscellaneous expenses, Popular major, Job Seeking Cost, Amount of student loan, CET Passage, Class ranking, Family financial assistance amount, Origin Level, Father's Occupation Level, Father's Education Level, Mother's Occupation Level

b Dependent Variable: Expected annual earnings

Table 4.19 Model Summary of Multiple Regression

R	R Square	Adjusted R Square
.367(a)	.135	.126

a Predictors: (Constant), Mother's Education Level, working experience, Amount of scholarship, Miscellaneous expenses, Popular major, Job Seeking Cost, Amount of student loan, CET Passage, Class ranking, Family financial assistant amount, Origin Level, Father's Occupation Level, Father's Education Level, Mother's Occupation Level

Table 4.18 shows the pilot regression model successfully passes the F test, which basically means the pilot regression estimate is good to apply to the whole population. The model is a good fit. Table 4.19 shows the R square of the pilot

regression is 0.135.

As to the individual predictors, not all of them are statistically significant. The t tests for independent variables failed 6 of the 14 predictors, including class ranking, student loan amount, miscellaneous expenses, mother's occupation level, father's education level and, mother's education level. 8 of the predictors are significant in this pilot regression. They are: popular major, College English Test passage, origin level, working experience, scholarship amount, job seeking cost, family financial assistance amount and, father's occupation level. The detailed results are presented by Table 4.20

Table 4.20 Coefficients and Co-linearity Estimates of Multiple Regression Model (a)

	Coef.	t	Sig.	Collinearity	
	B			Tolerance	VIF
(Constant)	10597.908	10.630	.000		
Popular major	4316.066	7.384	.000**	.924	1.083
CET Passage	1825.321	4.388	.000**	.896	1.116
Origin Level	2348.720	2.845	.003**	.813	1.230
working experience	-1952.855	-3.124	.002**	.952	1.051
Class ranking	332.471	1.053	.292	.912	1.096
Amount of scholarship	.288	2.941	.003**	.926	1.078
Amount of student loan	.124	1.815	.069	.941	1.063
Job Seeking Cost	.472	2.296	.022**	.951	1.051
Family financial assistance amount	.045	2.764	.006**	.862	1.133
Miscellaneous expenses	.008	.209	.834	.931	1.075
Father's Occupation Level	2053.093	2.652	.008**	.553	1.807
Mother's Occupation Level	-420.795	-.491	.624	.546	1.833
Father's Education Level	1129.483	1.261	.209	.652	1.534
Mother's Education Level	133.048	.177	.859	.684	1.462

a Dependent Variable: Expected annual earnings

**: significant at 5% level

In addition to the predictor coefficients, Table 4.20 also reports the collinearities among predictors. The "Collinearity Statistics Tolerance" is calculated as 1 minus R square for an independent variable when it is predicted by the other independent variables already included in the analysis. This statistic may be interpreted such that a variable with very low tolerance contributes little information to a model, and can cause computational problems. Almost all predictors' tolerance values are sufficiently large to hold their validity except father's occupation level, which has a fairly low tolerance of 0.553.

Since the pilot linear regress just throws everything into one regression equation regardless of the potential influential factors and colliniarities among predictors, the model could be imprecise. A stepwise method therefore is implemented to improve the accuracy of the model prediction.

When stepwise method is applied, independent variables can be entered or removed from the model depending on either the significance (probability) of the F value or the F value itself. All independent variables must pass the tolerance criterion to be entered in the equation, regardless of the entry method specified. Also, an independent variable is not entered if it would cause the tolerance of another variable already in the model to drop below the tolerance criterion. Table 4.21 is the result of the stepwise regression.

Table 4.21 Coefficients and Collinearity Estimates of Stepwise Regression [a]

Model		Coefficients	t	Sig.	Collinearity	
		B			Tolerance	VIF
1	(Constant)	14978.622	32.459	.000		
	Popular major	5491.660	9.432	.000	1.000	1.000
2	(Constant)	13047.719	23.278	.000		
	Popular major	4980.638	8.585	.000	.978	1.023
	CET Passage	2413.629	5.918	.000	.978	1.023
3	(Constant)	12361.232	21.543	.000		
	Popular major	4920.083	8.528	.000	.977	1.023
	CET Passage	2170.779	5.325	.000	.963	1.038
	Father's Occupation Level	2851.202	4.835	.000	.983	1.017
4	(Constant)	13161.282	21.366	.000		
	Popular major	4581.566	8.495	.000	.977	1.023
	CET Passage	1914.452	4.640	.000	.932	1.073
	Father's Occupation Level	3027.953	5.137	.000	.976	1.025
	working experience	-2186.950	-3.479	.001	.962	1.039
5	(Constant)	13070.439	21.278	.000		
	Popular major	4778.549	8.334	.000	.974	1.026
	CET Passage	1816.258	4.407	.000	.928	1.078
	Father's Occupation Level	2491.370	4.092	.000	.908	1.101
	working experience	-2222.927	-3.549	.000	.962	1.039
	Origin Level	2531.149	3.343	.001	.915	1.093
6	(Constant)	12804.392	20.687	.000		
	Popular major	4631.397	8.071	.000	.967	1.034
	CET Passage	1772.530	4.311	.000	.928	1.079
	Father's Occupation Level	2583.899	4.251	.000	.906	1.104
	working experience	-2139.922	-3.423	.001	.960	1.041
	Origin Level	2593.327	3.434	.001	.914	1.094
	Amount of scholarship	.285	2.984	.003	.984	1.016
7	(Constant)	12015.655	17.794	.000		
	Popular major	4371.617	7.547	.000	.944	1.060
	CET Passage	1723.080	4.198	.000	.925	1.081
	Father's Occupation Level	2270.682	3.688	.000	.876	1.140
	working experience	-2084.444	-3.342	.001	.959	1.042
	Origin Level	2556.295	3.393	.001	.914	1.094
	Amount of scholarship	.301	3.154	.002	.981	1.020
	Family financial assistance amount	.046	2.882	.004	.929	1.076

8	(Constant)	11630.186	16.743	.000		
	Popular major	4443.160	7.671	.000**	.941	1.063
	CET Passage	1730.704	4.224	.000**	.926	1.081
	Father's Occupation Level	2152.580	3.489	.001**	.871	1.148
	working experience	-2060.883	-3.309	.001**	.959	1.043
	Origin Level	2462.564	3.269	.001**	.911	1.097
	Amount of scholarship	.284	2.977	.003**	.975	1.026
	Family financial assistance amount	.044	2.758	.006**	.927	1.079
	Job Seeking Cost	.488	2.304	.021**	.976	1.024

a Dependent Variable: Expected annual earnings

**: significant at 5% level

The last section of Table 4.21 is the last step of the stepwise regression, e.g., the 8th model tried by the stepwise process. 8 variables are significant predictors to the dependent variable. They are: popular major, CET passage, father's occupational level, working experience, origin level, scholarship amount, family financial assistance amount and job seeking cost. Though the stepwise regression concurs the pilot regression's result on significant independent variables, it computes the coefficient slightly different and also yields different tolerance value for each variable by dropping the non-significant predictors. However, the stepwise method sacrifices a small amount of regression of the estimate model. Table 4.22 and 4.23 show the model fit summary the stepwise estimates equation, which comparing to Table 4.20 shrinks a little bit (0.126<0.125) on model regression.

Table 4.22 Model Fit of Stepwise Regression (j)

Model		Sum of Squares	df	F	Sig.
8	Regression	2.03E+10	8	25.330	.000(h)

	Residual	1.35E+11	1349		
	Total	1.55E+11	1357		

h Predictors: (Constant), Popular major, CET Passage, Father's Occupation Level, working experience, Origin Level, Amount of scholarship, Family financial assistant amount, Job Seeking Cost

j Dependent Variable: Expected annual earnings

Table 4.23 Model Summary of Stepwise Regression

Model	R	R^2	Adjusted R^2
1	.248(a)	.062	.061
2	.292(b)	.085	.084
3	.317(c)	.101	.099
4	.330(d)	.109	.106
5	.341(e)	.116	.113
6	.349(f)	.122	.118
7	.357(g)	.127	.123
8	.361(h)	.131	.125

a Predictors: (Constant), Popular major

b Predictors: (Constant), Popular major, CET Passage

c Predictors: (Constant), Popular major, CET Passage, Father's Occupation Level

d Predictors: (Constant), Popular major, CET Passage, Father's Occupation Level, working experience

e Predictors: (Constant), Popular major, CET Passage, Father's Occupation Level, working experience, Origin Level

f Predictors: (Constant), Popular major, CET Passage, Father's Occupation Level, working experience, Origin Level, Amount of scholarship

g Predictors: (Constant), Popular major, CET Passage, Father's Occupation Level, working experience, Origin Level, Amount of scholarship, Family financial assistant amount

h Predictors: (Constant), Popular major, CET Passage, Father's Occupation Level, working experience, Origin Level, Amount of scholarship, Family financial assistant amount, Job Seeking Cost

IV.3.1.2 Linear Regression with Selected Independent Variables

According to the results of the preceding pilot regression and stepwise regression, several significant predictors are identified, including: popular major, CET passage, father's occupational level, working experience, origin level,

scholarship amount, family financial assistance amount and job seeking cost. Yet, these significant predictors are either put into the linear equation without selection or selected by stepwise function of the statistical application automatically. Both methods could be either inaccurate or biased.

In this section, the independent variables are selected and dropped manually by the investigator according to certain criteria to improve the model fit and the accuracy of prediction.

Some family socioeconomic status characters such as parents' education level and occupation level are, in most occasions, latent variables, which indirectly impact the earnings through other factors. Therefore these latent SES factors might or might not be significant if they are put directly into the earnings equation. Considering the complexity and intangibleness of the latent family SES factors, predictors like father and mother's education level and occupation are dropped from the linear regression equation. Yet, these family SES factors are very important to the estimation of a person's earnings, and this is also the hypothesis in this study. They will be examined with more advanced econometric method in the latter part of this chapter. However, some other SES variables are considered as direct SES variables, which influence the earnings in a more tangible or measurable manner. Such direct SES variables include miscellaneous expenses, family financial assistance amount and, student loan amount are put into the

earnings estimate function. Other predictors like popular major, CET passage, working experience and job seeking cost, scholarship amount and, origin level, which are significant in both of the preceding regression, are kept in the earnings function. As a result, 10 independent variables are identified and selected into the earnings estimate equation. They are: Popular major (PopMaj), College English Test passage (CETpss), origin type level (OrgnLv), working experience (WkExpc), overall class ranking (Rank), scholarship amount (Schshipa), student loan amount (LoanAmnt), job seeking cost (JbSkCst), family financial assistance (Familyas), miscellaneous expenses (MisclExpns).

The new earnings estimate regression model has a good fit to the sample and could be well applied to the population. Table 4.24 and 4.25 give out the model summary of new regression. Obviously, the regression of the R squares is better than both of the previous model (0.126>0.125). That is, after manually selection of independent variables, the model is more powerful and accurate in earnings estimation.

Table 4.24 Model Fit of Manu Selection Model (b)

	Sum of Squares	df	F	Sig.
Regression	2.14E+10	10	21.201	.000(a)
Residual	1.40E+11	1389		
Total	1.62E+11	1399		

a Predictors: (Constant), Miscellaneous expenses, CET Passage, Amount of student loan, Job Seeking Cost, Amount of scholarship, Origin Level, working experience, Popular major, Class ranking, Family financial assistance amount

b Dependent Variable: Expected annual earnings, N=1400

Table 4.25 Model Summary of Manu Selection Model

R	R Square	Adjusted R Square
.364(a)	.132	.126

a Predictors: (Constant), Miscellaneous expenses, CET Passage, Amount of student loan, Job Seeking Cost, Amount of scholarship, Origin Level, working experience, Popular major, Class ranking, Family financial assistant amount

Though the model as a whole piece is better than the previous models, it doesn't necessarily guarantee each individual predictor is better than those in the previous models. The independent variable t test and collinearity test is conducted. The results are shown in Table 4.26.

Table 4.26 Linear Regression with Enter Process (a)

Independent variables (Predictors)	Unstandardized Coefficients	t	Sig	Collinearity Statistics	
	B			Toferance	VIF
(Constant)	10689.19	10.68	.00		
Popular major	4561.28	7.88	.00**	.92	1.09
CET Passage	2103.27	5.14	.00**	.91	1.10
Origin Level	3310.49	4.51	.00**	.96	1.05
working experience	-1747.61	-2.62	.00**	.96	1.04
Class ranking	281.19	.91	.36	.92	1.09
Amount of scholarship	.27	2.73	.01**	.93	1.08
Amount of student loan	.11	1.71	.09	.94	1.06
Job Seeking Cost	.49	2.45	.01**	.97	1.04
Family financial assistance	.05	3.39	.00**	.91	1.10
Miscellaneous expenses	.01	.29	.77	.94	1.07

a Dependent Variable: Expected annual earnings, N=1400

**: significant at 5% level

7 of the 10 predictors are tested to be significant in estimating earnings. They are: job seeking cost, family financial assistance amount, amount of scholarship, popular major, CET passage, origin level and, working experience. Comparing to the first stepwise regression result, the new regression model inherit all the significant predictors from the former model except father's occupation

level, which is dropped by the investigator.

Table 4.26 shows that 1400 cases are observed, which means these cases have their expected initial earnings reported. The regression passes the model F test. For the independent variable t tests, FATHSCHL, MOTHSCHL, MATRITYP, SCHLP and POPMAJ have very small p value, which means they are significant at 5% confidence level. However, MISCLEXP, OWNSHP, FATHOCLV, MOTHOCLV and ORGNTYPR don't pass the t test.

Let's take a look at the coefficients. WKEXPC, which stands for working experience, is tested to be significantly affecting the expected earnings. Yet, the negative coefficient looks confusing. WKEXPC is coded as 0 for no experience, 1 for having experience. Accordingly, the result is that those students with working experiences expect ¥1747.61 less than those who do not.

FAMILYAS, e.g., amount of family financial assistance has significant impact on initial earnings. The coefficient of the independent variable is 0.05, which indicates that one Yuan increase of family financial assistance yields 0.05 Yuan change in initial annual earnings. The propensity is the more financial assistance a student gets from the sibs and friends, the more he or she is expected to make in the future.

LOANAMNT, e.g., the amount of student loan, is a significant predictor for expected earnings in this regression model. The coefficient of the predictor is

0.11, which means one Yuan of change in student loan amount yields 0.11 Yuan change in the expected earnings per year, in the same direction. That is, the more a student borrows for his or her college education, the higher wage he or she is going to make in the future.

Class ranking (RANK) and miscellaneous expenses (MISCLEXP) fail to be significant predictors in this regression model, so is amount of student loan. Two direct family socioeconomic status variables including origin level and family financial assistance amount are significant; however, miscellaneous expenses failed the t test.

To examine the validity of the investigator's predictor selection and model building, a stepwise procedure is introduced again. All 10 independent variables are entered into the stepwise model. The results are shown in Table 4.27

Table 4.27 Model Fit of Linear Regression with Stepwise Process

Model		Sum of Squares	df	F	Sig.
1	Regression	1.10E+10	1	101.845	.000(a)
	Residual	1.51E+11	1398		
	Total	1.62E+11	1399		
2	Regression	1.62E+10	2	72.647	.000(b)
	Residual	1.46E+11	1397		
	Total	1.62E+11	1399		
3	Regression	1.73E+10	3	55.981	.000(c)
	Residual	1.44E+11	1396		
	Total	1.62E+11	1399		
4	Regression	1.85E+10	4	45.054	.000(d)
	Residual	1.43E+11	1395		
	Total	1.62E+11	1399		

5	Regression	1.95E+10	5	38.292	.000(e)
	Residual	1.42E+11	1394		
	Total	1.62E+11	1399		
6	Regression	2.04E+10	6	33.552	.000(f)
	Residual	1.41E+11	1393		
	Total	1.62E+11	1399		
7	Regression	2.10E+10	7	29.707	.000(g)
	Residual	1.41E+11	1392		
	Total	1.62E+11	1399		

a Predictors: (Constant), Popular major

b Predictors: (Constant), Popular major, CET Passage

c Predictors: (Constant), Popular major, CET Passage, Origin Level

d Predictors: (Constant), Popular major, CET Passage, Origin Level, Family financial assistance amount

e Predictors: (Constant), Popular major, CET Passage, Origin Level, Family financial assistance amount, working experience

f Predictors: (Constant), Popular major, CET Passage, Origin Level, Family financial assistance amount, working experience, Amount of scholarship

g Predictors: (Constant), Popular major, CET Passage, Origin Level, Family financial assistance amount, working experience, Amount of scholarship, Job Seeking Cost

The last section of Table 4.27 shows the model summary when the 7th, and also the last predictor is entered to the equation. Obviously, R square regression is not improved, if it's not decreased, which basically means the new earnings estimate equation is at least as good as it could be. The model fit has no room to improve even with the automatic selection by the statistic computer application. Table 4.28 gives out the independent variable coefficients and the corresponding correlations of each predictor with other ones.

Table 4.28 Linear Regression with Stepwise Process (a)

		Models						
		1	2	3	4	5	6	7
(Constant)	Unstd. Coef.	14768.054	12743.416	12444.909	11523.068	12271.730	11977.660	11545.621
	Sig.	.000	.000	.000	.000	.000	.000	.000
Popular major	Unstd. Coef.	5789.065	5210.999	5058.548	4769.704	4746.292	4583.952	4677.989

	Sig.	.000	.000	.000	.000	.000	.000	.000
CET Passage	Unstd. Coef.		2558.983	2382.495	2301.613	2084.052	2040.586	2041.670
	Sig.		.000	.000	.000	.000	.000	.000
Origin Level	Unstd. Coef.			3314.651	3131.702	3226.509	3309.494	3177.793
	Sig.			.000	.000	.000	.000	.000
Family financial assistant	Unstd. Coef.				.052	.052	.056	.052
	Sig.				.001	.001	.000	.001
working experience	Unstd. Coef.					-1963.703	-1878.188	-1855.406
	Sig.					.002	.002	.003
Amount of scholarship	Unstd. Coef.						.262	.266
	Sig.						.003	.005
Job Seeking Cost	Unstd. Coef.							.487
	Sig.							.015

a Dependent Variable: Expected annual earnings

The last section shows the final step of the stepwise selection process with the entry of the 7th and the last significant predictor. Though all 7 variables are significant, popular major has a fairly high level of collinearity with other predictors, which should be noticed. However, it's not high enough to be problematic.

In the regard of the collinearities among the independent variables, the correlation test is conducted as follows.

Table 4.29 Pearson Correlations of Stepwise Regression

	Expct anni earning	Pop major	CET Pass	Orign Levl	wkng expnc	Class ranking	scholar ship	studen t loan	Job Skng Cost	Family fincl assist	Miscell expens
Expct anni earnings		.26	.20	.15	-.11	-.03	.10	.06	.07	.14	.07
Pop major	.26		.16	.07	-.05	-.04	.09	.09	-.04	.16	.11
CET Pass	.20	.16		.11	-.17	-.17	.05	.02	.01	.09	.00
Orign Levl	.15	.07	.11		.02	-.06	-.03	-.10	.07	.09	.09
wkng expnc	-.11	-.05	-.17	.02		.00	-.06	-.09	-.01	-.02	.00

Class ranking	-.03	-.04	-.17	-.08	.00	[redacted]	-.21	-.01	-.07	.04	.04
scholarship	.10	.09	.05	-.03	-.08	-.21	[redacted]	.13	.06	-.05	.00
student loan	.06	.09	.02	-.10	-.09	-.01	.13	[redacted]	-.02	-.13	-.01
Job Skng Cost	.07	-.04	.01	.07	-.01	-.07	.06	-.02	[redacted]	.07	.12
Family fincll assist	.14	.16	.09	.09	-.02	.04	-.05	-.13	.07	[redacted]	.20
Miscell expens	.07	.11	.00	.09	.00	.04	.00	-.01	.12	.20	[redacted]
Sig. (1-tailed) at 5% level											
Expct annl earning	.	.00	.00	.00	.00	.12	.00	.02	.00	.00	.01
Pop major	.00**	.	.00	.00	.03	.07	.00	.00	.05	.00	.00
CET Pass	.00**	.00**	.	.00	.00	.00	.03	.22	.37	.00	.48
Origin Levl	.00**	.00**	.00**	.	.23	.01	.10	.00	.00	.00	.00
wkng expnc	.00**	.03**	.00**	.23	.	.48	.01	.00	.29	.21	.48
Class ranking	.12	.07	.00**	.01**	.48	.	.00	.30	.00	.08	.05
scholarship	.00**	.00**	.03**	.10	.01**	.00**	.	.00	.01	.03	.47
student loan	.02**	.00**	.22	.00**	.00**	.30	.00**	.	.26	.00	.32
Job Seeking Cost	.00**	.05**	.37	.00**	.29	.00**	.01**	.26	.	.01	.00
Family fincll assist	.00**	.00**	.00**	.00**	.21	.08	.03**	.00**	.01**	.	.00
Miscell expens	.01**	.00**	.48	.00**	.48	.05**	.47	.32	.00**	.00**	.

Table 4.29 manifestly shows that some independent variables are correlated to each other in a fairly high degree. Most of them are in fact significantly correlated. These correlations undermine the validity of the each individual predictor. It basically means a proportion of the error reduction of a predictor is shared with other predictors.

The lower section of Table 4.29 gives out the one-tailed significance of the correlations. For example, popular major significantly correlated with CET passage, origin level, working experience, class ranking, scholarship, student loan, job seeking cost, family financial assistance and miscellaneous expenses. Two sets of collinearities should be noticed: popular major and origin level are almost

significantly correlated with all other independent variables, except class ranking and scholarship separately. A reasonable explanation to this phenomenon is that latent family socioeconomic status factors influence those variables through popular major and origin level. More specifically, for instance, popular major is negatively correlated with working experience, which means popular major students tend to have no working experience. From another perspective, popular major is positively correlated with miscellaneous expenses, which means the reduction in error of earnings estimation yielded by popular major should be showed partly with the miscellaneous expenses of the student.

For origin level, on the one hand, it is positively correlated with job seeking cost, family financial assistance and miscellaneous expenses; on the other hand, it is negatively correlated with class ranking and student loan. The latent socioeconomic status factors are believed to embed in theses correlations. The relationships are quite intuitive. That is, when estimating the earnings, some predictors are overlapped on their effects, high SES students tend to have more miscellaneous expenses and family financial assistance in the meanwhile they also tend to from better origin (city or metro area) and spend more money on their job hunting in return to higher return in labor market. One fact should be clear here, which is somehow not quite intuitive. The variable of family financial assistance (FAMILYAS) refers to the monetary support or gift from sibs and friends to assist

or support the specific student's college study and campus life. Parents' financial support to the student is excluded. Generally speaking, people tend to think that poor students from low SES family have more family financial assistant than those high SES ones.

However, the real situation is in fact counter common sense. High SES students tend to possess more family financial assistance instead of low SES students. The rationale is: low SES students' families tend to have fewer wealthy sibs and friends who are able to support the students financially. Moreover, in the current social setting of China, it is not only the capability but also the willingness taking effect in this regard. A high SES household usually possess large amount of social resources, especially the heads of the households are usually possess a high level of occupation. In the scenario of China, it is very likely to be government officials or managers. The family sibs or friends will be willing to give monetary gift to the children when they go to college in return to their parents' favor in the past or future. This is so called "soft bribery", which is quite popular in China. Table 4.26 is the result of Kendall's tau_b nonparametric correlation test of family fiancial assistance (FAMILYAS), father's occupation type (FathOccTy) and mother's occupation type (MothOccTy). Family financial assistance is significantly correlated with father's occupation type, mother's occupation type, and both correlations are positive. Since the occupation type is coded in an

descending order, positive correlation with family financial assistance means higher level occupation type is associated with higher amount of family financial assistance. The higher-SES-high-FAMILYAS hypothesis is supported by the evidence.

Table 4.30 Correlations between FathOccTy, MothOccTy and FAMILYAS

		FathOccTy	MothOccTy	FAMILYAS
FathOccTy	Correlation Coefficient	1.000	.661(**)	.177(**)
	Sig. (1-tailed)	.	.000	.000
	N	9100	8903	5886
MothOccTy	Correlation Coefficient	.661(**)	1.000	.158(**)
	Sig. (1-tailed)	.000	.	.000
	N	8903	9602	6224

**: Correlation is significant at the 0.01 level (1-tailed).

In short, most of the correlations between the predictors in the new regression model shown in Table 4.30 are associated with latent family SES characteristics. It indicates that though latent SES variables are not in the estimate equation, they still impact earnings indirectly. Unfortunately, the ordinary linear regression technique is not able to identify the effect of the latent SES variables.

The observed cases in this regression is 1400, while several thousands of other cases are not examined simply because these cases do not have reported expected annual earnings, which implies that by that time, these individuals did not have confirmed job offers and income estimates. Could these missing cases tell a different story in the earnings prediction model? It's quite possible, since the size of the censored observations take almost 4/5 of the total sample.

As mentioned in the preceding part, the selection hypothesis is: the selection of whether a college graduate gets a job offer or not by the end of the survey is not random; the selectivity bias is due to family socioeconomic status. To examine this hypothesis, an advanced econometric model is needed to take the censored observation into account and correct the selectivity bias. Heckman two-stage consistent estimate method is therefore employed to achieve this objective.

IV.3.2 A Heckman Two-stage Consistent Estimator Model Analysis

First, an array of variables hypothesized to influence the initial earnings of the college graduates should be identified.

Variables POPMAJ, CETPSS, ORGNLV, WKEXPC, RANK, SCHSHIPA, LOANAMNT, JBSKCST, FAMILYAS and MISCLEXP are k vector of $[X_k]$, which means they are in regression equation (i) and; variable FATHOCCL, MOTHOCCL, FATHEDLV, MOTHEDLV and ORGNLV are j vector of $[Z_j]$, which means they are in selection equation (ii).

It should be noted that several variables in the data set are ordinal variables, which can not be used directly in the equation as selection instrumental variables,

dummy variables are needed. Some ordinal variables are recoded into boolean variables. For example, for "LOAN", 1 stands for having student loan, 0 stands for none. All other selection parameters are recoded based on the median of the descriptive analysis: 1 stands for equal or above median, 0 stands for below median. All selection instrumental variables are dichotomous; please refer to Table 4.3 for detailed recoding transformation.

The family socioeconomic status may include a lot of components such as parents' income, parents' education attainment, parents' occupation, family wealth, location of residency, and etc. There are several important family SES characteristics in this data set. As mentioned in the preceding part, the SES factors are divided into two categories: latent SES factor and direct SES factor, depending on the measurability. The latent SES factors in this study include: father's highest schooling level, mother's highest schooling level, FathOccTy and mother's occupation type. The direct SES variables include: origin type, student loan, miscellaneous expenses and family financial assistance amount. All latent SES variables are recoded dichotomously and put into the selection function, e.g., equation (ii), while all direct SES variables are put into the earnings estimate equation, e.g., equation (i). Origin type is put in both equations due to its unique characteristic and measurability.

Miscellaneous expenses are all the expenses other than food, lodging,

tuition and school fees. To some degree, it reflects the aggregate purchasing power and the overall financial situation of the student. Yet, the variance of the miscellaneous expenses is not only caused by the student's financial situation, it is also influenced by the specific individual's spending behavior pattern, or say, consumption custom. So it might have high co-linearity with consumption custom. Unfortunately, the data set doesn't provide the information about the individual's consumption custom and it is not possible to do the covariance analysis to control for the consumption custom.

Family financial assistance recorded the amount of financial assistance the student got from his or her relatives and friends, excluding the money from the parents. This variable is also controversial in reflecting the family SES background. On the one hand, student from a low-SES household could be short of financial resources to attend the college and need the more family financial assistance from the sibs, then yield the result of high family financial assistance. On the other hand, however in China, a special scenario might also happen. For instance, a student from a high-SES family, say, a government official family, the sibs or friends could also provide a lot of "soft money" to the student when they attend college, in return for their own goods or benefits. Hence, family financial assistance might not accurately reflect the family socioeconomic status, e.g., a low-family financial assistance value could refer to a low-SES background as well as a high-SES

background. In these regards, FAMILYAS and MISCLEXP need special attention when interpreting the result.

Table 4.31: Heckman Two-stage Method Estimates

Number of obs. = 6792, Censored obs.=5434, Uncensored obs.=1358
Wald chi2(11) =188.43, Prob > chi2=0.0000

	Coef.	z	P>|z|	[95% Conf.	Interval]
mills					
lambda	6574.525	3.26	0.001**	2622.238	10526.81
rho	0.57118				
sigma	11510.48				
lambda	6574.525				
POPMAJ	4291.724	7.38	0**	3.15E+03	5.43E+03
CETPSS	1797.173	4.34	0**	9.85E+02	2.61E+03
ORGNLV	920.233	0.88	0.377	-1.12E+03	2.96E+03
MKEXPC	-1915.505	-3.08	0.002**	-3.13E+03	-6.97E+02
RANK	371.242	1.18	0.237	-2.44E+02	9.87E+02
SCHSHIPA	0.287	2.93	0.003**	9.52E-02	4.78E-01
LOANAMNT	0.124	1.82	0.069	-9.77E-03	2.57E-01
JBSKCST	0.453	2.22	0.026**	5.36E-02	8.53E-01
FAMILYAS	0.047	2.87	0.004**	1.48E-02	7.84E-02
MISCLEXP	0.009	0.22	0.826	-6.97E-02	8.73E-02
_cons	2751.657	1.02	0.308	-2.54E+03	8.04E+03
select					
FATHOCCL	-0.22777	-4.82	0	-0.32032	-0.13522
MOTHOCCL	-0.11837	-2.37	0.018	-0.21617	-0.02058
FATHEDLV	-0.07404	-1.71	0.086	-0.15868	0.010606
MOTHEDLV	-0.12867	-2.72	0.007	-0.22136	-0.03598
ORGNLV	-0.27562	-6.02	0	-0.36536	-0.18586
_cons	-0.54061	-20.55	0	-0.60093	-0.49628

**: Sig. at 5% level.

In this Heckit model estimation, the uncensored observation number is 1358, which has a 42 observations reduction from the previous linear regressions due to the statistical application automatic data trimming function. The censored

cases number is 5434. Chi square value approaches to zero, which indicates the model fit is very good. For the selectivity bias test, the inverse mills ratio value, e.g., the lambda (λ) equals 3.26, which is larger than 1.96. It implies that the selectivity bias does exist in the current employment status, i.e., whether an expected annual earnings is reported in the sample. The null hypothesis is rejected, which means, family socioeconomic status differences are associated with the selectivity bias of job offers of the college graduates in this sample. In the other words, whether colleges graduate have job offer and report expected annual earnings by the end of survey is affected by the student's family socioeconomic status.

With the selectivity bias holding controlled by Heckman's two-stage method, the predictors to the expected earnings do have some differences with the previous regression model predictions. Table 4.32 gives out the detailed changes.

Table 4.32 Independent Variable Coefficient Comparison

	Pilot Regression		Stepwise Regression		Heckman Two-stage	
Lable	Coef.	Sig.	Coef.	Sig.	Coef.	Sig.
POPMAJ	4561.26	0**	4677.99↑	.00**	4291.72↓↓	0**
CHIPEB	2103.27	0**	2041.67↓	.00**	1797.174↓	0**
ORGNLV	3310.49	0**	3177.79↓	.00**	920.23↓↓	0.377
NKEXPC	-1747.6	0.005**	-1855.41↑	.00**	-1915.51↓↓	0.002**
SCHSHIPA	0.26675	0.006**	.266↓	.01**	0.2868↑↑	0.003**
JBSRCST	0.49436	0.015**	.487↓	.02**	0.45↓↓	0.028**
FAMILYAS	0.05396	0.001**	.052↓	.00**	0.047↓↓	0.004**

- The ↓and↑ symbols in the forth and sixth column indicate the changing direction of the coefficient comparing to the previous models.
- **: Sig. at 5% level.

By applying Heckman two-stage method, the origin type level (ORGNLV) becomes not significant anymore, while all other 6 variables continue to be significant from the previous regressions.

With current employment status selectivity bias controlled, there are six predictors are significant.

Popular major (POPMAJ), is significant in all models, and it has a very large coefficient in the equations. Controlling for selectivity bias of current employment status caused by family SES factors, the coefficient shrinks, which means popular major has collinearity with family SES variables. Interpreting in another manner, some of the predicting power or effect of popular major to the earnings is shared by family SES variables. When the family SES factors are taking into account, the effect size of popular major to earnings is reduced. The correlation test results in Table 4.33 support the explanation. Whether a student gets admitted to a popular major, is positively correlated with the values of the parents' occupation type and negatively correlated with parents' education level. To be remembered that, occupation and education level are coded in a descending order from high to low, e.g., a lower value means of lower occupation level and lower education level. The results show that students from high-SES families with highly educated and positioned parents tend to choose popular majors and tend to get higher pay after graduation. For origin type, it is negatively correlated with

popular major. Since origin type is also coded in a descending order, negative correlation with popular major implies that low-SES student tend to enrolled in non-popular majors.

Table 4.33 Spearman's Correlation Test of POPMAJ and SES characteristics

		FathOccTy	MothOccTy	FHSL	MHSL	OrgnTy
FathOccTy	Correlation Coefficient	1.000	.762(**)	.650(**)	.560(**)	-.606(**)
	Sig. (1-tailed)	.	.000	.000	.000	.000
	N	9100	8903	9026	8850	8958
MothOccTy	Correlation Coefficient	.762(**)	1.000	.557(**)	.634(**)	-.651(**)
	Sig. (1-tailed)	.000	.	.000	.000	.000
	N	8903	9602	9467	9475	9456
FHSL	Correlation Coefficient	.650(**)	.557(**)	1.000	.633(**)	.474(**)
	Sig. (1-tailed)	.000	.000	.	.000	.000
	N	9026	9467	9745	9545	9586
MHSL	Correlation Coefficient	.560(**)	.634(**)	.633(**)	1.000	.526(**)
	Sig. (1-tailed)	.000	.000	.000	.	.000
	N	8850	9475	9545	9611	9461
OrgnTy	Correlation Coefficient	-.606(**)	-.651(**)	.474(**)	.526(**)	1.000
	Sig. (1-tailed)	.000	.000	.000	.000	.
	N	8958	9456	9586	9461	9870
Popular major	Correlation Coefficient	.035(**)	.032(**)	-.041(**)	-.044(**)	-.031(**)
	Sig. (1-tailed)	.001	.001	.000	.000	.001
	N	8849	9344	9487	9350	9596

**: Correlation is significant at the 0.01 level (1-tailed).

In short, the popularity of major significantly affects a college student's expected annual earnings averagely by RMB￥4291.724, and this effect is exaggerated by RMB￥269.54 (4561.26-4291.72=269.54) by family SES factors. That is to say that if people look at whether a college student is enrolled in a popular major, it will be the primary factor to determine the student's earnings;

moreover, without any knowledge or consideration of the student's family socioeconomic background, the impact of the popular major will be larger than it supposed to be.

Origin type level (ORGNLV) is significant in both pilot regression and stepwise regression models. However, it becomes not significant anymore when selectivity bias control kicks in. This change indicates that origin type's impact on earnings estimate has high collinearity with other SES factors, especially those latent SES variables, which are not in the linear earnings estimate regression equation. Since latent SES variables are excluded from the linear regression equation, due to the high collinearity, origin type takes all the shared effects and becomes a significant predictor. Moreover, the coefficient of origin type level is also quite large in both pilot regression and stepwise regression equation and drops dramatically (3310.49→920.23) in the Heckman two-stage model equation, which also indicates, its high collinearity with the latent SES variables including parents' education and occupation level. From this perspective, the impact of family socioeconomic status to the college graduate's expected annual earnings are supported by the result with a fairly large magnitude.

Working experience (WKEXPC) has a large coefficient and, continues to be significant after selectivity bias is controlled. Additionally, the coefficient even increased (in absolute value) after the correction of selectivity bias. This indicates

that students with working experience during or prior to their college years earns less than who do not. Specifically to this study, the earnings gap is 1915.51 Yuan/year. There are two possible explanations to this phenomenon: (1) for those non-traditional students, i.e., attend college after working in labor market for a certain period of time, they might have lower-SES and academic attributes than traditional student and consequently earn less then traditional student, even though the non-traditional ones have the same education level; (2) for those traditional student, some of them, most likely from low-SES families, have to work part timely during their college years to support their study, they also make less future earnings comparing to their high-SES counterparts due to their low-SES attributes. Explanation (1) is supported by the nonparametric test.

Table 4.34 Median Comparison Traditional Student vs. Student with Working Experience

	FathOccTy	MothOccTy	FHSL	MHSL	Origin type	Class ranking	Miscell expns	CET Pass
N	8814	9062	9195	9069	9286	9235	6389	9273
Median	3.00	2.00	4.00	5.00	2.00	2.00	4000.00	1.00
Chi-Square	.715	12.756	4.465	1.158	2.620	10.387	6.506	18.634
df	1	1	1	1	1	1	1	1
Asymp. Sig.	.398	.000	.035	.282	.106	.001	.011	.000

a Grouping Variable: working experience before college

Table 4.34 shows that mother's occupation type, father's highest schooling level, class ranking miscellaneous expenses and CET passage are significantly different between students, who have working experience before college and those who don't. Shown in Table 4.35 the means of traditional students tend to be better

off on family SES and academic than students with working experience before college.

Table 4.35 Means Descriptive

		MothOccTy	FHSL	CET Pass	Rank	Miscell expns
Traditional Student	Valid	6710	6835	6917	6879	6157
	Missing	393	268	186	224	2946
Mean		3.35	4.16	1.08	1.97	6791.97
Working Experience	Valid	352	350	356	358	241
	Missing	14	6	10	10	125
Mean		3.74	3.85	.85	1.78	5996.34

Therefore explanation (1), is somewhat supported by these evidences. Though traditional students perform better academically, however, lower academic performance does not cause the lower expected earnings, considering RANK is non-significant in all models. Therefore, family SES background attributes to the differences.

The patterns are clear: non-traditional students are more likely come from low-SES families, have lower academic preparation for college education, perform poorer in college and get lower pay after graduation.

However, for explanation (2), the SES and academic tendency of working student and non-working student are quite ambiguous. Table 4.36 and 4.37 show mix propensities, hence, the explanation (2) is lack of evidence to prove.

Table 4.36 Median Comparison Working Student vs. Non-working Student

	FathOccTy	MothOccTy	FHSL	MHSL	Origin type	Class ranking	Miscel. Expns.	CET Pass
N	8263	8701	8826	8709	8937	8871	8156	8909
Median	3.00	2.00	4.00	5.00	2.00	2.00	4800.00	1.00

Chi-Square	.292	6.156	.229	.050	2.510	13.833	4.641	25.360
df	1	1	1	1	1	1	1	1
Asymp. Sig.	.589	.013	.632	.823	.113	.000	.031	.000

a Grouping Variable: working experience before college

Table 4.37 Means Descriptive

		MothOccTy	CET Pass	Rank	Miscel Expns
Non-Working Student	Valid	6331	6475	6449	4471
	Missing	299	155	181	2159
Mean		3.31	1.11	1.95	6666.26
Working Student	Valid	2370	2434	2422	1684
	Missing	93	29	41	779
Mean		3.46	1.01	2.04	6598.65

Therefore, working experience (WKEXPC) is correlated with expected earnings negatively through family socioeconomic status and regardless of college attainment.

The same story with popular major happens to College English Test passage (CETPSS). CETPSS has the same trend of change with popular major (POPMAJ) across the three models. It continues to be significant after selectivity bias is controlled; however, its coefficient shrinks by 306.1 Yuan (2103.27-1797.17). That indicates the collinearity with SES factors, i.e., the SES variables share part of their effect on earnings estimates with CET passage. Generally speaking the CET passage has sound impact on college students' earnings, yet the actual impact of CET passage sole is not as large as it appears to be if the employment selectivity bias causing by the family SES is taken into consideration. A college graduate, who passes CET Band 6 earns RMB

￥1797.17/year more than those who pass CET 4, and those who failed in CET test earns 1797.17 Yuan/year less than those who pass CET Band 4.

Family financial assistance (FAMILYAS) is the last significant predictor in both linear regressions and Heckman two-stage model. As mentioned in the earlier part, there are two possible explanations for this result: (a) students from low-SES families need more financial assistance from sibs and friends, and tend to work harder academically, and then yield a higher expected earnings; (b) students from high-SES families are more likely to obtain financial assistance in forms of awards or gifts from the sibs and parents' friends in return to their parents' favor. Obviously, the first explanation is invalid because of the non-significance of RANK. That is, college performance has nothing to do with expected earnings.

Table 4.38Non-parametric Correlations Analysis

		FathOccTy	MothOccTy
FathOccTy	Correlation Coefficient	1.000	.661(**)
	Sig. (1-tailed)	.	.000
	N	9100	8903
MothOccTy	Correlation Coefficient	.661(**)	1.000
	Sig. (1-tailed)	.000	.
	N	8903	9602
FAMILYAS	Correlation Coefficient	.177(**)	.158(**)
	Sig. (1-tailed)	.000	.000
	N	5883	6224

** Correlation is significant at the 0.01 level (1-tailed).

Table 4.38 is the non-parametric correlations analysis result of Family financial assistance amount with parents' occupation type. Since Family financial

assistance amount is a scale variable, and FathOccTy and mother's occupation type are categorical variables, Kendall's tau-b correlation test is applied, which is good for both scale and ordinal variables. The test shows significant results between family financial assistance and parents' occupation type, and they are both positively correlated. Considering the coding sequence of the value of occupation type is ascending from lowest level to the highest level, students' with higher occupation type parents' get more family financial assistance, and vice versa. Therefore, the second possible explanation is supported by the result i.e., FAMILYAS positively correlated with expected initial annual earnings because of the family socioeconomic status. In the meanwhile, since the coefficient shrinks by introducing Heckman two-stage method to control for selectivity bias, as explained in the preceding part, the shrinking of coefficient indicates the collinearity with latent SES variables.

Amount of scholarship (SCHSHIPA) is also a significant predictor in all models. However, after controlling the selectivity bias of current employment status, the coefficient of SCHSHIPA increases. It indicates that by taking out the family socioeconomics status's effect on the current employment status, the real impact of scholarship amount to the annual earnings is actually larger then people normally think. In other words, family SES doesn't have collinearity with scholarship amount; moreover, its effect on employment status reduces the

scholarship's effect on earnings. If all family SES factors held constant, the more scholarship a student is awarded, the higher wage he or she makes. Table 4.39 provides the evidence to this explanation. Amount of scholarship has no significant correlations with most of the major SES variables.

Table 4.39 Kendall's tau_b Correlation test of SCHLP and SES characteristics

		FathOccTy	MothOccTy	FHSL	MHSL	SchalpAmnt
FathOccTy	Corr. Coef.	1.000	.661(**)	-.534(**)	-.457(**)	-.011
	Sig. (1-tailed)	.	.000	.000	.000	.129
	N	9100	8903	9026	8850	6205
MothOccTy	Corr. Coef.	.661(**)	1.000	-.458(**)	-.531(**)	-.014
	Sig. (1-tailed)	.000	.	.000	.000	.071
	N	8903	9602	9467	9475	6583
FHSL	Corr. Coef.	-.534(**)	-.458(**)	1.000	.546(**)	.012
	Sig. (1-tailed)	.000	.000	.	.000	.098
	N	9026	9467	9745	9545	6623
MHSL	Corr. Coef.	-.457(**)	-.531(**)	.546(**)	1.000	-.016(*)
	Sig. (1-tailed)	.000	.000	.000	.	.032
	N	8850	9475	9545	9611	6586

**: Correlation is significant at the 0.01 level (1-tailed).

*: Correlation is significant at the 0.05 level (1-tailed).

Four other predictors are non-significant, including overall class ranking (RANK), miscellaneous expenses (MISCLEXP), ownership of institution (OWNSHP), mother's occupation level (MOTHOCLV) and origin type (ORGNTYPR).

IV.4 Alternative Models

IV.4.1 Alternative Model 1: Heckman Two-stage Estimates with Institutional Characteristics

In the preceding parts, 6 of the independent variables have been identified as significant predictors to college graduates' expected initial annual earnings, and they are believed to correlate with earnings through the latent effect of family socioeconomic status. Besides these 6 independent variables, are there any other factors, alternatively also impact the earnings, but is independent to family socioeconomic status? An alternative model is built to test such factors.

Exogenous factors like ownership of the institution and matriculation type are not examined in the previous models, which might also have impact on the college graduates' initial earnings. In China, though the majority of the higher education institutions body is under public ownership, as mentioned in the previous chapter, since two decades ago, private or semi-private higher education institutions have emerged to become a more and more important component of the higher education system. These so called "Min Ban", or people run institutions recruit students normally with lower academic preparation and charged them higher tuition and fees. In fact, these private or semi-private colleges and schools are in general less comparative and reputable than their public counterpart. Might the ownership of the institution contribute to the variation of the earnings of their graduates? In the meanwhile, matriculation type of the college students could also

influence the graduates' earnings for the same reason. Almost began in the same time with the emerging of multi-ownership of the tertiary institution, the matriculation type of college enrollments were diversified. As mentioned in the earlier chapter, public institutions were allowed to charge more tuition and fees to recruit freshman with lower College Entrance Examination (CEE) score beyond the national college enrollment annual quota designated by the ministry of education. Those students enter the college with out-of-plan matriculation status or go to non-public institutions pay more tuition and fees to the schools. For sure, these students are financially different from their public counterpart, which implies the potential differences in family social economic status. Hence, both ownership of the institution and matriculation type of the student could impact the earnings.

Table 4.40 Alternative Model 1 with Heckman Two-stage Method

Number of obs.= 6790, Censored obs.=5434, Uncensored obs.=1356 Wald chi²(13)= 188.88, Prob > chi²= 0.0000					
	Coef.	z	P>|z|	[95% Conf.	Interval]
mills					
lambda	6676.499	3.29	0.001**	2700.916	10652.08
rho	0.57796				
sigma	11551.86				
lambda	6676.499				
POPNAJ	4207.236	7.21	0**	3063.115	5351.358
CETPSS	1772.2	4.27	0**	958.4574	2585.942
NKEXPC	-1849.5	-2.97	0.003**	-3071	-628.006
SCHSHIPA	0.300266	3.06	0.002**	0.107813	0.492719
JBSKCST	0.452577	2.22	0.026**	0.053062	0.852092
FAMILYAS	0.046436	2.86	0.004**	0.014587	0.078285
ORGNLV	907.7473	0.87	0.384	-1136.41	2951.901

RANK	368.9055	1.17	0.24	-246.732	984.5431
LOANAMNT	0.122949	1.81	0.071	-0.01053	0.256424
MISCLEXP	0.007897	0.2	0.844	-0.07057	0.086368
MATRITYP	3924.328	1.35	0.177	-1778.62	9627.277
OWNSHP	402.8587	0.1	0.922	-7659.09	8464.812
_cons	-1611.34	-0.28	0.78	-12895.8	9673.069
select					
FATEDUCL	-0.22727	-4.81	0	-0.31983	-0.13471
MOTHEDUCL	-0.11822	-2.37	0.018	-0.21602	-0.02042
FATHEDLV	-0.07293	-1.69	0.091	-0.1576	0.011732
MOTHEDLV	-0.12856	-2.72	0.007	-0.22125	-0.03586
ORGNLV	-0.27527	-6.01	0	-0.36503	-0.1855
_cons	-0.55029	-20.6	0	-0.60265	-0.49792

**: Sig. at 5% level

Table 4.40 shows the results of alternative model 1. Both matriculation type and ownership of institution are not significant, which means in terms of estimating earnings, both variables do not have significant impact on the dependent variable when family SES factors are controlled. The result is not surprising in fact, because the independent variable of OWNSHP and MATRITYP themselves are composite of various elements. Both high SES and low SES students could end up with enrolling in a non-public institution or an out-of-plan matriculation status. The tricky thing here is interaction between the academic preparation and family SES.

A good performing student could be forced to enroll in a non-public institution or with an out-of-plan matriculation status if he or she does not have a good family SES. In China, college administrations highly depend on the College

Entrance Examination (CEE), however, the bottom line or minimum requirement of the CEE score is not strictly fixed, especially when the candidates' score are around the bottom line. In general, most of the colleges require more qualified candidates than they actually allowed recruiting under nation quota from College Entrance Examination Commission (CEEC), which is the ultimate gatekeeper of all CEE exam takers. Consequently, the college will inform those low CEE score candidates, especially whose score are around the bottom line, to choose if they want to be matriculated with out-of-plan status. If the institution possesses good reputation and prestige, even out-of-plan matriculation is attractive enough for a lot of students, since once they get the admission, there will be no difference with other students on their diploma when they eventually graduate. Therefore, students with barely good enough CEE score and higher SES are more likely to choose enrolled with out-of-plan matriculation status than waiting to be recruited by a lower prestige institution.

On the other hand, low SES students are more likely to have low CEE score. Figure 4.1 and 4.2 shows clear trends that the fewer years of schooling or the lower level occupation a student's parents possess, the lower he or she is likely to score in CEE test.

Figure 4.1 Mean CEE score by Parents' Years of Schooling

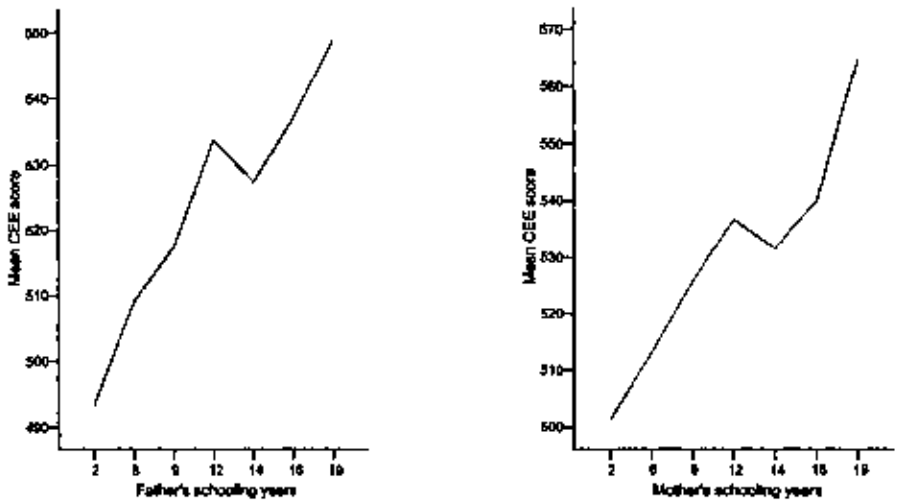
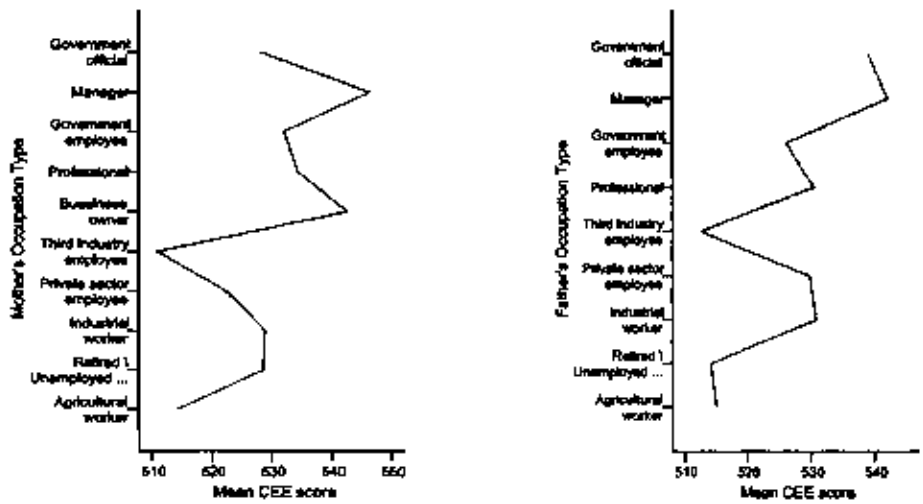


Figure 4.2 Mean CEE score by Parents' Occupation Type



Therefore in general, low SES students tend to enroll in non-public institutions or with an out-of-plan matriculation status, yet when the students' CEE score close to the bottom line of the admission requirement, higher SES students are more likely to get the admissions. That is, family socioeconomic status' impact on matriculation type and ownership of institution could be very complicated and with no constant direction. Consequently, MATRITY and OWNSHP are not

significant in alternative model 1.

IV.4.2 Alternative Model 2: Gender-Specific Estimates

Though exogenous variables like MATRITY and OWNSHP are not significant, variable like gender, however, could have significant impact on earnings and should be considered as an alternative explanation. By 2000, male to female ration of the national population is about 1.439 (MOE, 2004), and the male to female ration of this sample is about 1.375. The discrepancy is reasonable. Because in the sample, there is 4.68% of the observation with missing value of gender. Due to the limitation of the data, the alternative model 2 is based on this data set. The model is based on the previous one; the difference is that gender is added in to the earnings equation. Table 4.41 shows the result.

Table 4.41 Heckman Two-stage Method Estimates with Gender

Number of obs.=6787, Censored obs.=5434, Uncensored obs.=1353 Wald chi2(12) = 199.02, Prob. > chi2=0.000					
	Coef.	z	P>|z|	[95% Conf.	Interval]
mills					
lambda	7105.144	3.50	0.00**	3.12E+03	1.11E+04
rho	0.606				
sigma	11719.653				
lambda	7105.144				
POPMAJ	4129.773	7.04	0**	2.98E+03	5.28E+03
CETPSS	1910.500	4.6	0**	1.10E+03	2.73E+03
WKEXPC	-1932.700	-3.12	0.002**	-3.15E+03	-7.18E+02
SCHSHIPA	0.285	2.92	0.003**	9.40E-02	4.77E-01

LOANAMNT	0.140	1.97	0.048**	1.03E-03	2.78E-01
JBEXCST	0.446	2.2	0.028**	4.84E-02	8.44E-01
FAMILYAS	0.048	2.94	0.003**	1.58E-02	7.93E-02
SEX	1549.801	2.52	0.012**	3.46E+02	2.75E+03
ORGNLV	997.934	0.95	0.341	-1.05E+03	3.05E+03
MISCLEXP	0.010	0.24	0.809	-6.86E-02	8.79E-02
RANK	266.442	0.84	0.399	-3.53E+02	8.86E+02
_cons	1139.970	0.41	0.681	-4.29E+03	6.57E+03
select					
FATHOCCL	-0.230	-4.87	0	-3.23E-01	-1.38E-01
MOTHOCCL	-0.113	-2.26	0.024	-2.11E-01	-1.50E-02
FATHEDLV	-0.076	-1.75	0.08	-1.61E-01	8.94E-03
MOTHEDLV	-0.128	-2.71	0.007	-2.21E-01	-3.55E-02
ORGNLV	-0.277	-6.03	0	-3.66E-01	-1.87E-01
_cons	-0.550	-20.6	0	-6.03E-01	-4.98E-01

**: Sig. at 5% level

The Heckman two-stage estimator model gives a significant inverse mill's ratio value (3.50 > 1.96), indicates the existence of selectivity bias causing by the latent SES factors. The overall model fit is also very good (Prob. > Chi^2 =0), which means the model fits the sample and is valid to be used on the whole population. Sex is shown significant in the alternative model 2, with a coefficient of 1549.801. That is, a male college graduate statistically earns 1549.801 Yuan each year more than a female. This result is quite intuitive and reasonable, and it concurs the previous studies done in other countries like the United States (Taubman & Wales, 1974).

Considering the significant impact of gender on the earnings estimation, a gender specified test is conducted to examine the gender's impact on earnings and

other predictors throughout.

The sample is divided into two sub-groups by gender, and then Heckman two-stage model is applied to each of the sub-group to estimate the same set of independent variables' correlation with earnings.

Table 4.42 Heckman Two-stage Method Estimates on Sub-groups by Gender

	Number of obs=4162, Censored obs=3228, Uncensored obs=934, Wald chi2(11) =127.33, Prob > chi2=0.0000			Number of obs = 2589, Censored obs = 2172, Uncensored obs=417, Wald chi2(11) =78.94, Prob > chi2=0.0000		
	Male			Female		
	Coef.	z	P>|z|	Coef.	z	P>|z|
mills						
lambda	7366.686	2.86	0.004**	8714.371	2.4	0.016
rho	0.60439			0.75291		
sigma	12188.6			11574.23		
POPMAJ	4077.277	5.48	0**	4327.04	4.75	0**
CNTPES	2852.419	5.31	0**	229.1256	0.37	0.714
ORGNLV	1232.415	0.85	0.394	70.87265	0.05	0.961
WKSKPC	-2086.43	-2.68	0.007**	-1893.93	-1.93	0.054
RANK	399.6663	1.05	0.296	-199.884	-0.36	0.717
SCHSHIPA	0.211707	1.73	0.083	0.417456	2.68	0.007**
LOANAMNT	0.118549	1.5	0.134	0.325511	1.84	0.066
JBSECST	0.245294	0.88	0.377	0.650622	2.36	0.018**
FAMILYAS	0.049002	2.34	0.019**	0.041804	1.7	0.09
MISCLEXP	-0.01443	-0.3	0.763	0.091413	1.26	0.207
_cons	2021.458	0.62	0.538	120.6764	0.02	0.982
select						
FATHOCCL	-0.24745	-4.16	0	-0.17475	-2.25	0.025
MOTHOCCL	-0.12637	-1.94	0.052	-0.05705	-0.72	0.471
FATHEDLV	-0.06011	-0.9	0.369	-0.19565	-2.39	0.017
MOTHEDLV	-0.14741	-2.46	0.014	-0.07311	-0.98	0.327
ORGNLV	-0.30442	-4.94	0	-0.18523	-2.62	0.009
_cons	-0.50017	-17.17	0	-0.72789	-16.05	0

**: Sig. at 5% level

Table 4.42 shows the results from both sub-groups. With family SES factors and employment status selectivity bias controlled, male graduates and female graduates appear to have quite different characteristics in terms of earnings estimation.

Table 4.42 shows that, for a male college graduate wage earner, a popular major will bring him 4077.3 Yuan more per year, every level of College English Test passage will bring him 2852.4Yuan more per year, a traditional student status (no working experience before graduate) will bring him 2086.4 Yuan more per year, and finally, every Yuan of family financial assistance will bring him 5 cents more earnings per year. For a female college graduate, a popular major will let her earn 4327 Yuan more per year, every Yuan she spends in job hunting and gets from scholarship is believed to bring her 0.65 Yuan and 0.42 Yuan more wage earnings per year.

Table 4.42 also shows that popular major is the only independent variable that is significant for both the male and female sub-samples. The other significant factors are different for the two sub-samples. For example, CETPSS, WKEXPC, and FAMILYAS are significant for males but not females; and SCHSHIPA and JBSKCST are significant for females but not males. In other words, in the current high-skill labor market in China, male college graduates will benefit a lot more from college education if he is from a high SES family and

masters English better. On the other hand, for female college graduates, her scholarship awarded is quite important for her earnings, so is how much she spends on job hunting. In short, the determinants of initial earnings are quite gender specific; and the labor market in China appears to operate different for male and female college graduates.

It should be noted that the number of significant predictor declines dramatically, i.e., drops from 6 to 4 for male and 6 to 3 for female. Why the significance of the independent variables change so dramatically? Let's take a look at the correlations between gender and the 9 independent variables in all previous models. Since some of the variables are ordinal variable, nonparametric correlation test is conducted and partial correlation analysis is also conducted for the reference.

Table 4.43 Nonparametric Correlation and Partial Correlation Table

		Gender	
		Nonparametric	Partial
Expected annual earnings	Correlation	-.017	.055**
Popular major	Correlation	.169**	.114**
CET Passage	Correlation	-.109**	-.130**
Origin Level	Correlation	-.133**	-.076**
working experience	Correlation	.082**	.005
Class ranking	Correlation	.182**	.145**
Amount of scholarship	Correlation	-.094**	-.038
Amount of student loan	Correlation	.115**	.102**
Job Seeking Cost	Correlation	-.032**	-.008
Family financial assistance	Correlation	-.039**	-.004

- **. Correlation is significant at the 0.05 level (1-tailed).
- Control variables in Partial correlation analysis= Father's Occupation Level & Mother's Occupation Level & Father's

highest schooling level & Mother's highest schooling level

Clearly, in Table 4.43 all 9 independent variables have significant correlation with gender based on nonparametric test. Since the estimated equations based on male and female sub-samples removes the influence of gender, the reduction in the number of significant predictors in the gender specific alternative model equations is reasonable.

Table 4.44 Coefficients Comparison Aggregate vs. Gender Specific Model

	Coef.		
	Male	All	Female
POWAJ	4077.277**	4129.773**	4327.04**
CETPSS	2652.419**	1910.500**	229.1256
WKEXPC	-2086.43**	-1932.700**	-1893.93
SCHSHIPA	0.211707	0.285**	0.417456**
JBHRCHT	0.245294	0.446**	0.650622**
FAMILYAS	0.049002**	0.048**	0.041804

**: significant at 5% level

Table 4.44 reveals that the estimated coefficients of the aggregated model (alternative model 1) are the weighted average of those coefficients in the male and female earnings equations (alternative model 2). Thus the estimated coefficients of the aggregate sample lie between those of the gender sub-samples.

IV.4.3 Alternative Model3: Heckman Two-stage Estimates with Ethnicity

Though race, as mentioned in the chapter III, is not an issue in China's case,

ethnicity can not be omitted. China has 56 ethnicities including Han. By 2000, 8.41% of the total population was minorities (NBSC, 2003). The minority observation in this sample is 10.7%, which is a little bit higher than the national figure. Since the Census 2000 data is the latest available national data set and there is a 3-year gap between the Census 2000 and this survey, the discrepancy in between is not regarded as a major inconsistency. Alternative model 3 examines the ethnicity's influence on initial earnings. Based on alternative model 2, ethnicity is added in this model to estimate the earnings.

Table 4.45 Heckman Two-stage Method Estimates with Ethnicity

				Number of obs=6347, Censored obs=5064,	
				Uncensored obs=1283, Wald chi2(12) =190.66, Prob > chi2=0.0000	
	Coef.	z	P>|z|	[95% Conf.	Interval]
mills					
lambda	6722.13	3.51	0**	2972.526	10471.73
rho	0.58404				
sigma	11509.63				
lambda	6722.13				
POPMAJ	4234.568	7.09	0**	3063.423	5405.712
CETFSS	1960.07	4.63	0**	1131.152	2788.988
ORGNLV	815.355	0.78	0.436	-1236.27	2866.983
WKEXPC	-2000.98	-3.15	0.002**	-3244.81	-757.159
RANK	355.498	1.1	0.27	-276.516	987.5116
SCRSHIPA	0.269173	2.73	0.006**	0.075684	0.462663
LOANAMNT	0.161803	2.25	0.024**	0.020979	0.302626
JBSKCST	0.053569	0.24	0.807	-0.37696	0.484095
FAMILYAS	0.049558	2.98	0.003**	0.016958	0.082158
MISCLEXP	0.0123	0.28	0.776	-0.07246	0.097055
SEX	1429.903	2.28	0.022**	203.1631	2656.643
STENT	-1107.59	-0.86	0.389	-3629.63	1414.451

_cons	2767.767	0.96	0.335	-2862.68	8398.214
select					
FATHOCCL	-0.24437	-5.06	0	-0.33909	-0.14964
MOTHOCCL	-0.11442	-2.22	0.026	-0.21543	-0.01342
FATHEDLV	-0.12109	-2.28	0.023	-0.22536	-0.01691
MOTHEDLV	-0.15042	-3.13	0.002	-0.24467	-0.05618
ORGNLV	-0.26711	-5.59	0	-0.3608	-0.17342
ETHNY	-0.04882	-0.6	0.549	-0.20838	0.110742
_cons	-0.49371	-6.11	0	-0.65215	-0.33528

**: significant at 5% level

Table 4.45 shows the result of the alternative model 3. Ethnicity as an independent variable in the earnings equation is not significant, which means in earnings equation, ethnicity does not have a significant impact on the initial earnings. However, the Inverse Mill's Ratio (Lambda) of the selection equation is significant ($z = 3.51 > 1.96$) with ethnicity as a selection variable. Ethnicity has indirect impacts on earnings. From alternative model 2 to alternative model 3, the only change in the selection equation is adding ethnicity as a latent family SES variable. Therefore, all differences between these two models should attribute to selectivity bias causing by ethnicity. Table 4.46 shows the detailed differences.

Table 4.46 Coefficients Comparison Alternative Model 3 vs. Alternative Model 2

	Alternative Model 2		Alternative Model 3	
	Coef.	P>|z|	Coef.	P>|z|
POPMAJ	4129.773	0**	4234.568	0**
CHTPSS	1910.500	0**	1960.07	0**
ORGNLV	997.934	0.95	815.355	0.44
WKEXPC	-1932.700	0.002**	-2000.98	0.002**
RANK	266.442	0.399	355.498	0.27
SCHSHIPA	0.285	0.003**	0.269173	0.006**
LOANAMNT	0.140	0.048**	0.161803	0.024**

JBSKCST	0.446	0.028**	0.053569	0.81
FAMILYAS	0.048	0.003**	0.049558	0.003**
MISCLNEP	0.01	0.809	0.0123	0.78
SEX	1549.601	0.012**	1429.903	0.022**
ETHNY			-1107.59	0.39

**: significant at 5% level

By adding ethnicity as a latent SES variable in the selection equation, job seeking cost (JBSKCST) becomes no longer significant in alternative model 3 comparing to alternative model 2. And the coefficients of other 7 significant predictors also change accordingly. All these changes ascribe to ethnicity's indirect impact on initial earnings. Therefore, the evidence is clear. Although ethnicity as an earnings predictor is not significant in the earnings equation, it impacts the earnings as a latent SES factor embedded in the process of college education and labor market selection.

Though the determinants of initial earnings in the labor market for college graduates are identified and quantified by this investigation, the reason why there is gender and ethnicity based differences is not adequately explained by the models.

IV. 5 Determinant of Employment Status —A Probit Analysis

Several significant independent variables were identified by the preceding analyses and the family socioeconomic status' impact on initial earnings was also captured by the models with Heckman two-stage methods. However, the mechanism of how the family SES factors influence the college graduates' initial earnings remains unclear. Multiple reasons might jointly yield the result, and the mechanism could be very complicated. One assumption is: family SES factors impact a college graduate's initial earnings through the process of employment, e.g., whether the student gets a job offer. That is, the likelihood of getting a job offer affects a college graduate's decision of what kind of job offer he or she will take (i.e., low pay or high pay job). During the process of employment decision making, family socioeconomic status factors and other factors, in turn, influence the propensity to get a certain wage-level job offer and then, the initial earnings.

To estimate the impact of these factors on initial earnings of this assumption, a probit analysis is employed. Since the variable of current employment status (CES) has two values: 1 and 0 represent "employed" and "unemployed" separately. There is no truncation selection problem. Table 4.47 shows the result of the probit estimates.

Table 4.47 Probit Estimates of Likelihood of Employment

Number of obs = 5230, LR $chi^2(16)$ = 209.43, Prob > chi^2=0.0000
Log likelihood = -3389.8231, Pseudo R^2=0.0300

CES	Coef.	z	P>|z|	[95% Conf.	Interval]
FATHEDLV	-0.108	-2.17	0.03**	-2.05E-01	-1.02E-02

CETPSS	-0.206	-7.7	0**	-2.58E-01	-1.53E-01
WKEXPC	-0.196	-4.87	0**	-2.75E-01	-1.17E-01
POPMAJ	0.150	3.84	0**	7.37E-02	2.27E-01
SEX	0.167	4.28	0**	9.04E-02	2.44E-01
MATRITYP	0.434	2.72	0.007**	1.21E-01	7.46E-01
OWNSHP	1.106	4.74	0**	6.49E-01	1.56E+00
FTHNY	-0.139	-1.7	0.09	-3.00E-01	2.17E-02
FATHOCCL	-0.053	-1.14	0.256	-1.45E-01	3.86E-02
MOTHOCCL	0.052	1.05	0.291	-4.45E-02	1.48E-01
MOTHEDLV	0.030	0.66	0.509	-5.99E-02	1.21E-01
ORGNLV	-0.083	-1.87	0.061	-1.70E-01	3.96E-03
RANK	-0.022	-0.96	0.338	-6.64E-02	2.28E-02
SCHLP	-0.012	-0.3	0.767	-9.37E-02	6.91E-02
LOAN	0.074	1.45	0.147	-2.61E-02	1.75E-01
JBEXCST	0.000	-0.47	0.641	-3.03E-05	1.87E-05
_cons	-0.950	-4.1	0	-1.40E+00	-4.96E-01

Dependent variable is CES

**: significant at 5% level

Shown in Table 4.47, there are seven significant independent variables in the probit model, which means these seven predictors determine the likelihood of a college graduate getting a job offer when he or she graduates. More specifically, popular major (POPMAJ), gender (SEX), matriculation type (MATRITYP) and ownership of the institution (OWNSHP) have positive coefficients while father's education level (FATHEDLV), CET passage (CETPSS) and working experience (WKEXPC) have negative coefficients. The coefficient for popular major is 0.150, which indicates that a popular major results in a 0.15 standard deviation increase in the predicted probit index. The coefficient for working experience is -0.196, which indicates that previous working experience decrease the likelihood of

getting a job offer by 0.196 standard deviation in the predicted probit index. Accordingly, the other coefficients for the predictors are interpreted in the same way.

Obviously, a male graduate from a public institution recruited under national recruiting plan with a popular major has higher probability to get a job offer, however the passage of College English Test (CET) decrease the probability of employment, so is the previous working experience. Most of the significant predictors coincide with the previous earnings estimates in terms of their impacts. However, for CET passage, the propensity of getting a job offer is totally different from the earnings prediction in the previous parts. One possibility to interpret this inconsistency of the propensity is that: during the process of employment and job hunting, a high level of CET passage yield a high self-expectation for the college graduates in job hunting, which consequently results in a lower employment ratio comparing to that of those low level CET passers. Table 4.48 provides the evidence to support this explanation.

Table 4.48 CET Passage by Current employment status Crosstabulation

CET Passage	Current employment status		Employed / Non-employed Ratio
	Non-employed	Employed	
Fail \ No score	1019	1226	1.20314
Band 4	1712	2830	1.653037
Band 6	1713	1170	0.683012
Total	4444	5226	1.175968

Shown in Table 4.48, the employed to non-employed ratio for Band 6

passage is only 0.68, which is only 41.3% of the ratio for Band 4 passers (1.653) and 56.8% of the ratio comparing to those do not pass the CET test. The propensity is very clear that high level CET passage is correlated with low employment rate.

Another possible explanation is that the high CET passers are very likely to continue their higher education at graduate level after graduation, since a Band 6 passage is a requirement for most of the graduate admission in China. Unfortunately, due to the limitation of the data set, this evidence of this assumption is not available.

Father's Education Level (FATHEDLV) also has negative correlation with current employment status, which seems quite counter-intuitive. The rational for this negative relationship is quite similar to that of CET passage. Table 4.49 provides the evidence.

Table 4.49 Father's Highest Schooling Level by Current Employment Status Crosstabulation

		Current employment status		Employed / Non-employed
		Non-employed	Employed	Ratio
Father's highest schooling level	Graduate	106	102	0.962
	undergraduate	710	670	0.944
	Associate	652	589	0.903
	High school	1208	1569	1.299
	Junior high	1070	1337	1.250
	elementary	585	813	1.390
	Illiterate or semi-illiterate	103	133	1.291
Total		4434	5213	1.176

The propensity shown in Table 4.49 is manifest: college graduate with a

low-educated father is more likely to accept a job offer then their high-patrilineal-educated counter part. To be remembered that, the median of father's education level for this sample is high school and the 1:1 turning point of the employed / non-employed ratio cuts right through high school level. The possible rational behind this phenomenon could be that more educated father transmits higher educational and occupational aspiration to the offspring and tends to encourage the next generation to pursue higher degree after graduation or to spend more time on seeking high level occupations. College graduates from families with high patrilineal education level face less financial pressure to work immediately and therefore the likelihood of employment declines accordingly.

The other significant predictors are quite intuitive. Working experience has negative correlation with employment status for the similar reason provided in the previous earnings analyses. Popular major, gender, ownership of the institution and matriculation type have positive correlation with employment status also for the similar reasons explained in the previous earnings analyses. Note that, most of such significant predictors as working experience, matriculation type, ownership of the institution, popular major, are correlated with family socioeconomic status or it is SES variable itself like father's education level.

To conclude, various factors including SES factors influence college graduates' initial earnings. The process of employment is one important channel

that family socioeconomic status transmits its impact from generation to generation in terms of initial earnings. The magnitude of the effect is measured and quantified via both direct and indirect means by this study, yet the other explanations or mechanisms are not provided by this study and are subject to further research.

CHAPTER V

SUMMARY OF FINDINGS AND DISCUSSION

This chapter summarizes the findings of this study based on the four key research questions: (1) Does higher education experience affect college graduates' initial earnings? (2) Is there any significant difference in initial earnings for college students from different socioeconomic backgrounds? How different are they? (3) Does the effect of family socioeconomic status on college graduates' initial earnings operate through higher education? (4) Are factors that influence initial earnings different for male graduates and female graduates?

V.1 Higher Education Experience Does Affect College Graduates' Initial Earnings

Higher education experience has several measures, among them three kinds of attainments are specified and estimated in this study. College English Test (CET) passage, merit-based scholarship and overall class ranking are examined. Both CET

passage and scholarship are found to be significant predictors to college graduates' initial earnings in the pilot multiple regression analysis, stepwise regression and Heckman two-stage models, while overall class ranking is not found any significance in any of the models. These results indicate that some, not all, measures of higher education experience have impact on the graduates' initial earnings.

The impact of CET passage on the initial earnings is quite substantial. The result of alternative model 3 reveals that, with selectivity bias corrected, holding others constant including gender and ethnicity, every level of CET passage will bring 1960.07Yuan more to a college graduate's pocket every year, which is about 9.3% of the mean total annual earnings.

The merit-based scholarship's impact on initial earnings is also quite significant. According to this investigation, every Yuan of scholarship a student gets during his or her school years will bring 0.27 Yuan more to his or her initial annual earnings.

All these evidences reveal the strong impact of higher education experience on initial earnings. Students with better college performance earn more than their colleagues, and the earnings gaps are quite substantial. The benefit of the quality of college education is realized in the labor market. From the perspective of Human Capital Theory, the quality of the skilled labor is significantly valued in China's high skill labor market nowadays. This study provides the first quantitative estimate of

higher education experience's contribution to individual's initial earnings with the correction of selectivity bias.

However, overall class ranking does not show any significance in any of the analysis in this study. There are several reasons that may explain such result. First, the self-reporting ranking data is not very reliable. Students may exaggerate their college performance. Second, the quartile ranking could vary substantially from college to college. A third quartile ranked student in a prestigious college could be ranked first quartile in an ordinary college. The third possibility is that the overall ranking may not actually have impact on initial earning at all. Due to the limitation of data, this investigation is not able to go further beyond these explanations on this item. The argument is subject to future research with more information.

V.2 The Difference in Initial Earnings for College Students From Different Socioeconomic Backgrounds is Significant

Socioeconomic status has been found significant in terms of its impact on one's earnings in the studies of developed countries such as the United States. This study uses advanced econometric method to examine the impact of SES on initial earnings in China.

This study finds that family socioeconomic status does influence college

graduates' initial earnings in China, by applying Heckman two-stage method both direct and indirect SES factors are examined. The comparison of non-selection-bias-controlling multiple regression and two-stage model reveals the impact of the latent SES factors is also significant.

The most profound finding of this study is using indirect approach to identify and quantify the impact of latent SES factor on initial earnings. This study finds that latent SES factors including parents' education level, occupation level and origin influence college graduates' initial earnings indirectly through non-SES factors, for instance popular major, working experience and scholarship. When the self-selection bias of family SES is controlled, the size and significance of the coefficient of other independent variables change. It is fair to say that family socioeconomic status permeate through almost every aspect of college education and employment process and influence an individual's initial earnings.

The way by which the SES background affects earnings is also examined by this study. Though the embedded mechanism could be quite complicated and ascribe to multiple reasons, one important reason is revealed by this study. Family SES background influences the initial earnings through the employment process. That is family SES first influence the college students' college attainment, e.g., the outcome of college education, which in turn influence the self-expectation of a certain level of job position and influence the decision making when the students graduates. High

SES graduates have less financial pressure yet more social/family resources to search high pay job and, tend to accept a job offer unhurriedly, while low SES graduates tend to accept available job offers more quickly due to their family financial pressure and self-selection. Through such channel family SES's influence infiltrates into the initial earnings.

One interesting phenomenon found in this study of employment is that though higher level of CET passage brings more earnings to the college graduates and high SES students tend to pass higher level CET, however, the employment ratio of high CET passage students are lower than low CET passage students. A possible reason is that high CET passage students also tend to pursue higher degree beyond college, invest more in their human capital to development better career in the future. Therefore, high CET passage yields high earnings but low employment ratio. This explanation has to be verified by further research.

V.3 The Effect of Family Socioeconomic Status on College Graduates' Initial Earnings Operates Through Higher Education

This study finds evidence to support the statement that family SES operate through higher education to influence initial earnings. Popular major and CET passage are found to be significant determinants of initial earnings, and both

predictors have significant correlation with family SES. The above findings suggest a logical chain of connections between SES, higher education and earnings.

College graduates from high SES families are more likely to enter a prestigious college under the national recruitment plan, study in a popular major, and have a high probability to pass CET Band 4 or even Band 6; while those from a low SES families are more likely to enter a less competitive institution with an out-of-plan matriculation, study in a non-popular major and have a low probability to pass the CET test. At graduation, high SES students are more likely to get a higher-paying job and end up with higher initial earnings, while the low SES students are more likely to end up with lower initial earnings. The already better off students tend to harvest more benefit from the process of college education, yet the already worse off students tend to derive less from higher education.

Thus family SES influences individual initial earnings through higher education. However, whether higher education amplifies or mitigates the earnings gap remains unsolved due to the limitation of the data set. Since the survey does not carry information before the college years and does not provide controlling groups of high school graduated workers, the initial earnings gap (if the college graduates would not have attended college and have entered the labor market directly after high school graduation) is not available.

V.4 Factors That Influence Initial Earnings are Different For Male Graduates and Female Graduates

Gender specific analyses find the earnings equations for male and female college graduates are quite different. Popular major is the only independent variable that is significant for both the male and female sub-samples, while all other predictors are different. In the current high-skill labor market in China, male college graduates will benefit a lot more from college education if they are from a high SES family and master English better. On the other hand, for female college graduates, their scholarship awarded is quite important for the earnings, so is how much they spend on job hunting. In short, the determinants of initial earnings are quite gender specific; and the labor market in China appears to operate differently for male and female college graduates.

Though the earnings equations have been estimated by this study and the predictors are identified for each of the gender group, the mechanism behind this pattern is unclear. Since this study only examine the earnings factors from the perspective of an individual, the macro-economic setting is not assessed. Apart from schooling, personal, and family characteristics, the examination of some other equally important elements of earnings are not conducted in this study such as labor

market preferences or discrimination, labor supply and demand and national economic growth. Without the assessment of these important components, the explanation of why male and female college graduates have different earnings equations is not feasible. This gap is to be filled by the future research.

The contribution of this study on the gender issue is to capture the different characteristics of earnings equations for male and female college graduates and provide a quantifiable base line plus reference for future studies.

V.5 Discussion

By exploring the responses to the four key research questions, this study finds the significant impact of family socioeconomic status on college graduates' higher education experience; and through this impact SES influences initial earnings in both direct and indirect means. The process of employment plays an important role as a medium to transmit the influence the family SES to initial earnings. Other factors like school characteristics and personal characteristics are also found to have significant impacts on initial earnings. Family socioeconomic status has broad influence in many aspects that determine the college graduates' initial earnings.

Similar study has not been conducted in China before. By adopting more advanced econometric technique, this investigation portraits the first image of the family SES's influence on college graduates' initial earnings in a quantitative manner. It provides an understanding of the current situation of the high skilled labor market in China after a serial of higher education reforms and national economic policy adjustments. This finding may help policy makers have a quantifiable perception of the relationship between family SES and returns to higher education and facilitate educational policy making in the future. Ultimately, this finding could help the policy makers make more pro-disadvantaged policies and lead the ongoing higher education reform to a more egalitarian direction. For example, the finding suggests that more egalitarian college admission and financial aid policies should pay more attention on students' family socioeconomic background. Higher academic standard may need to be applied on students from high SES background in the same institution to make the admission more equal. In the meantime, lower academic standard may be applied to low SES students for merit-based scholarship to give these disadvantaged students equal opportunity to offset the impact of low family SES.

This study explores earnings difference between male and female graduates. According to the sample, without controlling for other factors, male and female graduates have about the same initial earnings. Referring to Table 4.2, mean initial

annual earnings for female and male graduates are 21133.56 Yuan and 21144.39 Yuan separately; the difference is only about 10 Yuan per year. However, according to the alternative model 2, after controlling for SES and other factors, male graduates have an advantage of 1549.80Yuan (refer to Table 4.41) per year over female. The reason why male and female graduates in the sample have equal mean initial annual earnings is because the female graduates tend to have higher SES background in the sample, i.e., female graduates' mean parental schooling years and median patrilineal occupation level tend to be higher then male graduates (refer to Table 4.2). This finding may provide helpful information for the future policy making in such areas as education, employment and social welfare to promote the equal opportunity for female.

For example, different earnings equations derived from this study could point out the direction for future affirmative action policies in the job placement. The different earnings equations could guide the policy makers to make policies or laws to balance the specific earning determinants for different genders to make the average wage and employment opportunity more equal for female. Also, the finding of 1549.80Yuan annual earnings gap between genders (when family SES is controlled) provides strong evidence to support the different salary standards for female and male. Instead of a seemingly egalitarian unified salary standard for both genders, gender specific salary scales could indeed equalize female's earnings

with that of their male counterpart.

The finding of ethnicity's indirect impact on initial earnings could also draw the attentions of the public and the policy makers. Since the previous study did not correct for self-selection bias, and therefore was not able to find the significant impact of ethnicity's on initial earnings. The finding of indirect impact of ethnicity on earnings in this study could provide a rationale for an affirmative action policy for minority college graduates in working place.

The findings of this study also illustrate that the value of human capital has been realized in the current Chinese labor market for college graduates, higher education experience does make different in terms of earnings. This may facilitate the macro-level policy making by providing more accurate and quantifiable measures, especially in higher education financial reform, national man power strategic adjustment and the development of non-government post-secondary education.

V.6 Limitations

Though this is the first study in China to examine the influence of family SES on initial earnings with correction for selectivity bias, several limitations of

this study could be improved in the future. First, due to the data limitation, college graduates as the only targeting group could not illustrate the situation of the whole spectrum of Chinese working force. Second, due to the lack of information on other cohort, assessment of whether higher education amplifies or mitigates the earnings difference causing by family SES is not applicable. It makes this study impossible to issue any precise policy recommendation to improve the inequality in higher education and employment. Third, information on non-cognitive abilities is not available in the data, which prevents this study from differentiating the effect of ability bias of the individuals. Fourth, macro analysis such as the supply-demand chain and labor market preferences are not conducted in this study, which could help to explaining some findings of this study in greater details and broader context. Fifth, the information of current employment status of this sample needs special attention. College graduates are quite different from ordinary job seekers in the labor market, in terms of the timing and likelihood of getting a job. Though graduates might not have a job offer by the end of the survey, they were very likely to get employed soon.

In addition, all data of this sample is self-reported, which means the data might have distortion and the distribution might not be perfectly random. Though the survey designer (IEE-PKU) tried to avoid the normality problem, the process of questionnaire distribution and collection could still be problematic for several

practical reasons. However, this is still the best most current data available.

Some information such as non-cognitive attributes, family wealth, parental income and non-college-educated cohort are not collected by this survey. If these data were available in the future, the study could be more precise. For instance, the ability bias could be correct for the college entrance and employment; whether higher education exaggerates or mitigates the earnings gap could be identified; family SES's influence could be quite different when family wealth is taking into account. Additionally, if panel data were available, a comparative analysis over time will derive more useful information especially the trend of socioeconomic development.

Therefore, several survey designs could be improved in future data collection. First, instead of judgment-stratified sampling scheme, a random sampling scheme could better represent the characteristics of the population. More specifically, at institution level, the participant institutions should be randomly selected from all the public institutions under the Ministry of Education. After the institutions are located, questionnaires should be distributed randomly among students. This random sampling scheme will avoid the data distortion causing by applying arbitrary sampling ratio on different sample stratums. Second, the questionnaire should involve questions about students' family income and wealth information. Though direct questions such as: "How much do your parents earn?" or "What is

the total value of household asset?" are not appropriate, indirect questions like number of cars owned or internal area of the house could provide useful information about family wealth. Third, questions about non-cognitive ability should be involved to control for ability bias. Fourth, more indicators of college performance should be used to better estimate the impact of college education experience on earnings.

BIBLIOGROPHY

Alexander, K. L. & Eckland, B. K. (1977). High School Context and College Selectivity: Institutional Constraints in Educational Stratification. *Social Forces*, Vol.56, No.1 (Sep., 1977), 1660188.

Becker, G.S. & Chiswick, B.R. (1966). Education and the Distribution of Earnings. *The American Economic Review*, Vol. 56, No. 1/2 (Mar., 1966), 358-369

Becker, G.S. (1993). *Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education (3rd ed.)*. Chicago, Illinois: University of Chicago Press.

Behrman, J.R.; Taubman, P. (1986). Birth Order, Schooling, and Earnings. *Journal of Labor Economics*, Vol. 4, No. 3, Part 2: The Family and the Distribution of Economic Rewards (Jul., 1986), S121-S145.

Bian, Y.J. & Logan, J.R. (1996). Market Transition and the Persistence of Power: The Changing Stratification System in Urban China. *American Sociological Review*, Vol. 61, No.5 (Oct., 1996), 739-758.

Campbell, R.T. & Henretta, J.C. (1980). Status Claims and Status Attainment: The Determinants of Financial Well-Being. *The American Journal of Sociology*, Vol. 86, No. 3 (Nov., 1980), 618-629.

Carazzini, A. J.; Dugan, D.J.; Grabowski, H.G. (1972). Determinants and Distributional Aspects of Enrollment in U.S. Higher Education, *The Journal of Human Resources*, Vol.7, No.1 (Winter, 1972), 39-59

Carnoy, M. & H. Levin. (1985). *Schooling and Work in the Democratic State.* Stanford, CA: Stanford University Press.

Carnoy, M. (1995). Race Earnings Differentials. In Carnoy, M. (ed.), *International Encyclopedia of Economic of Education* (2nd ed.). New York: Elsevier Science Ltd.

Corazzini, A.J. (1967). When Should Vocational Education Begin? *The Journal of Human Resources*, Vol. 2, 41-50.

Corman, H. (2003). The Effects of State Policies, Individual Characteristics, Family Characteristics, and Neighbourhood Characteristics on Grade Repetition in the United States. *Economic of Education Review,* Vol.22 (2003), 409-420.

Cotter, D.A.; Hermsen, J.M.; Vanneman, R. (1999). Systems of Gender, Race, and Clas Inequality: Multilevel Analysis. *Social Forces*, Vol.78, No.2 (Dec., 1999), 433-460.

Denison, E.F. (1962). *The Sources of Economic Growth in the United States and the Alternatives Before Us.* New York: Committee for Economic Development.

Duncan, B. & Featherman, D.L. (1972). *Socioeconomic background and achievement.* New York: Seminar Press.

Fagerlind, I. & Saha, L. (1989). *Education and national development: A comparative perspective.* Oxford. Pergamon Press.

Fleisher, B.M.; Li, H.Z.; Li, S.; Wang, X.J. (Dec., 2004). Sorting, Selection, and Transformation of the Return to College Education in China. IZA Discussion Paper No. 1446, (Dec., 2004).

Griliches, G. (1977). Estimating the Returns to schooling: Some Econometric Problems. *Econometrica*, Vol. 45, 1-22.

Haider, S.J. (2001). Earnings Instability and Earnings Inequality of Males in the United States: 1967-1991. *Journal of Labor Economics*, Vol.19, No.4 (Oct., 2001), 799-836.

Hanushek, E. A. and Quigley, J. M. (Jun., 1985). Life-Cycle Earning Capacity and the OJT Investment Model. *International Economic Review*, Vol.26, No.2 (Jun., 1985), 365-385.

Hauser, R.M. & Sewell, W.H. (Jul., 1986). Family Effects in Simple Models of Education, Occupational Status, and Earnings: Findings from the Wisconsin and Kalamazoo Studies, *Journal of Labor economics*, Vol.4, No.3, Part2: The Family and the Distribution of Economic Rewards (Jul., 1986), S83-S115.

Hauser, R.M. & Warren, J.R. (1997). Socioeconomic Indexes fro Occupations: A Review, Update, and Critique, *Sociological Methodology*, Vol.27 (1997), 177-298.

Hennretta, J.C. & Campbell, R.T. (Mar., 1978). Net Worth as an Aspect of Status, The American Journal of Sociology, Vol.83, No.5 (Mar., 1978), 1204-1223.

Hackman, J.J. (1974). Shadow Wages, Market Wages and Labor Supply, *Econometrica,* Vol. 42, 679-693.

Heckman, J.J. (1979). Sample Selection Bias a Specification Error, *Economietrica,* Vol. 47, No.1, (Jan., 1972), 153-162.

Hotchkiss, L. & Borow, H. (1996). Sociological perspective on work and career development. In Brown, D. & Brooks, L. (Eds.), *Career Choice and Development* (3rd ed.)(pp. 281-334). San Francisco, CA: Jossey-Bass.

Iyanaga, S. & Kawada, E. (Eds.). (1980) Encyclopedic Dictionary of Mathematics. Cambridge, MA: MIT Press.

Jie, Z.G. (2004) Family Socio-economic Status and Possibility of Students' Employment, *Journal of Beijing Normal University (Social Science Edition)*, Vol.183, No.3 (2004).

Johnson, E.N. & Chow, G.C. (1997). Rates of Return to Schooling in China. *Pacific Economic Review*, Vol.2, No.2, 101-113.

Kenny, L.W.; Lee, L.F.; Maddala, G.S.; Trost, R.P. (1979). Returns to College Education: An Investigation of Self-Selection Bias Based on the Project Talent Data. *International Economic Review*, Vol.20, No.3, (Oct., 1979), 775-789.

La Rocque, N. & Jacobsen, V. (Jan., 2000). *A Market and regulatory Survey of Private Education in China.* (McKinsey working report, 2000).

Lazear, E. (1976). Age, Experience, and Wage Growth. *The American Economic Review*, Vol.66, No.4 (Sep., 1976), 548-558.

Lee, L. (1978). Unionism and Wage Rates: A Simultaneous Equations Model with Qualitative and Limited Dependent Variables. *International Economic Review*, 19(2), 415-433.

Lee, L. (1982). Some Approaches to the Correction of Selectivity Bias. *Review of Economic Studies*, XLIX, 355-372.

Levin, H.M. (1995). Work and Education. In Carnoy, M. (ed.), *International Encyclopedia of Economic of Education* (2nd ed.). New York: Elsevier Science Ltd.

Li, H.Z. (2003). Economic Transition and Returns to Education in China. *Economic of Education Review*, Vol.22, (2003), 317-328.

Lindsay, P. & Know, W.E. (1984). Continuity and Change in Work Values Among Young Adults: A longitudinal Study, *The American Journal of Sociology*, Vol.89, No.4, (Jan., 1984), 918-931.

Liu, J. (Jul., 2003). Analysis of Some Financing Reform Problems in China's Higher Education, *Shanghai Statistics*, March, 2003.

Liu, X.Y. (2004). *The Current Social Mobility in China*. The China Academy of Social Science Press.

Liu, Z. (1998). Earnings, Education, and Economic Reforms in Urban China. *Economic Development and Cultural Change,* Vol. 46, No.4, 687-725.

McMahon, W.W. (1987). Externalities of Education. In Psacharopoulos, G. (ed.), *Economics of Education: Research and studies,* pp. 133-137. Oxford: Pergamon.

Min, W. (2002). Economic Transition and Higher Education Reform in China (Note for Presentation). http://www.tc.edu/centers/coce/pdf_files/EconTransitionandHEReform.pdf

Mincer, J. (1958.) Investment in Human Capital and Personal Income Distribution. *The Journal of Political Economy,* Vol. 66, No. 4 (Aug., 1958), 281-302.

Ministry of Education, China. (2000). A Glimpse of Education Development of China, 2000. http://www.edu.cn/20011219/3014655_1.shtml#18

Ministry of Education, China. (2004). China Education Statistics Bulletin, 2003. May 27th, 2004. http://www.moe.edu.cn/edoas/website18/info5515.htm

Murphy, K. & Welch, F. (May, 1989). Earnings s for College Graduates: Recent Growth and Possible Explanations, *Educational Researcher,* Vol.18, No.4 (May, 1989), 17-26.

National Bureau of Statistics People's Republic of China. (03/28/2002) .Major Figures of the 2000 Population Census (No.1). http://www.sfpc.gov.cn/03/28/2002

National Center for Education Statistics (NCES). (1988). National Longitudinal Study of 1988. Retrieved February 01, 2004, from NCES Web site: http://www.nces.edu.gov

Rumberger, R.W. (1987). The impact of surplus schooling on productivity and earnings. *Journal of Human Resources*, 22 (1), 24-50.

Rumberger, R.W. (Mar., 1983). The Influence of Family Background on Education, Earnings, and Wealth. *Social Forces*, Vol.61, No.3, 755-773.

Rytina, S. (1992). Scaling the Intergenerational Continuity of Occupation: Is Occupational Inheritance Ascriptive After All? *The American Journal of Sociology*, Vol.97, No.6 (May, 1992), 1658-1688.

Schultz, T.W. (1961). Investment in Human Capital. *The American Economic Review*, Vol. 51, No. 1 (Mar., 1961), 1-17.

Schultz, T.W. (1963). *The Economic Value of Education.* New York. Columbia University Press.

Schultz, T.W. (1981). *Investing in People: The Economics of Population Quality.* Los Angeles. University of California Press.

Sewell, W.H. & Hauser, R.M. (1975). *Education, occupation, and earnings: Achievement in the early career.* New York: Academic Press.

Sewell, W.H.; Haller, A.O.; Portes, A. (Feb., 1969). The Educational and Early Occupational Attainment Process, *American Sociological Review*, Vol.34, No.1 (Feb., 1969), 82-92.

Smith, H.L. & Powell, B. (1990). Great Expectations: Variations in Income Expectations Among College Seniors. *Sociology of Education*, Vol.63, No.3 (Jul., 1990), 194-207.

Solomon, L.C. & Fagnano, C.L. (1995). Benefits of Education. In Carnoy, M. (ed.), *International Encyclopedia of Economic of Education* (2nd ed.). New York: Elsevier Science Ltd.

Solomon, L.C. (1981). New findings on the links between college education and work. *Higher Education*, 10 (6), 615-48.

Solow, R. M. (1965). Human Capital Review. *The Journal of Political Economy*, Vol.73, No.5. (Oct., 1965), 552-553.

Swanson, A.D. & King, R.A. (1991). *School Finance: Its Economics and Polities*. New York. Longman.

Taubman, P & Wales, T. (1974). Higher Education and Earnings. *McGraw-Hill Book Company*.

Taubman, P. & Wales, T. (1974). *Higher Education and Earnings: college as an Investmentand a Screening Device*. Report prepared for the Carnegie Commission on Higher Education and the National Bureau of Economic Research, General Series 101. New York: McGraw-Hill.

Tsang, M. (2000). Education and National Development in China since 1949: Oscillating Policies and Enduring Dilemmas. In Lau C. and et al. (ed.), *China Review 2000*. The Chinese University Press. Hong Kong, pp580-618

The Royal Swedish Academy of Sciences. (2000). Advanced information on the Bank of Sweden Prize in Economic Sciences in Memory of Alfred Nobel, 2000: The Scientific Contributions of James Heckman and Daniel McFadden. Stockholm, Sweden.

Vella, F. (1998). Estimating Models with Sample Selection Bias: A Survey. *The Journal of Human Resources*, Vol. 33,No.1(Winter, 1998), pp127-169.

Velloso, J. (1995). Income Distribution and Education. In Carnoy, M. (ed.), *International Encyclopedia of Economic of Education* (2nd ed.). New York: Elsevier Science Ltd.

Wanner, R.A. & Lewis, L.L. (1982). Trends in Education and Earnings, 1950-70: A Structural Analysis. *Social Forces*, Vol.61, No.2, (Dec., 1982), 436-455.

Warren, J.R.; Sheridan, J.T.; Hauser, R.M. (2002). Occupational Stratification across the Life Course: Evidence from the Wisconsin Longitudinal Study. *American Sociological Review*, Vol.67, No.3 (Jun., 2002), 432-455.

Wen, D.M. & Huang, L. (2004) The gender differences of graduate employment: a re-analysis. To be published by the Peking University Press.

Willis, Robert J. & Rosen, Sherwin. (1979). Education and Self-Selection, *The Journal of Political Economy*, Vol.87, No.5, Part2: Education and Income Distribution (Oct., 1979), S7-S36.

World Bank (2000). Higher Education in Developing Countries: Peril and Promise. Washington, DC: World Bank.

Wu, X.G. (2002). Work Units and Income Inequality The Effect of Market Transition in Urban China. *Social Forces,* Vol. 80, No.3 (Mar., 2002), 0037-7732.

Xun, G.Y.(Mar., 2004)The Influence and Countermeasure of the Disparity Between the Income Impressed by the unequal Opportunity of Higher Education, *Journal of Qujing Teachers College,* Vol.23, No.2, 2004.

Yu, D.H. & Lu, S. (2001). Higher Education Expansion's Effect to Income Equality in China, *Journal of Education and Economics*, No.1, 2001.

Zhao, Y.Z. (2000). Family SES's Influence to the Possibility of Higher Education, *Youth Study*, No. 3, 2000.

APPENDIX

Appendix I: Survey of the Willingness on Higher Education Institution Gradnates' Occupational Choice and Employment, 2003. (Original Chinese version)

高校毕业生择业行为与意愿调查问卷

亲爱的同学：

您好！首先祝贺您顺利毕业，并祝愿您找到理想的工作。

为更好地研究高等教育改革发展及毕业生求职状况，我们组织了这次调查。本调查只用于汇总统计分析，不会对您个人及所在学校带来任何不利影响。

为便于及时向您反馈本调查的有关统计结果，如相关专业同学的就业率、起薪状况、工作适应状况等，也请您能留下您的 Email 地址。谢谢！

教育部“十五”课题“高等教育规模扩展与劳动力市场”课题组

填答方式：

请在相应序号上打“√”，或在“___”中填答简要文字和数字。

第一部分：基本信息

1、您将长期使用的 Email 地址是：_____

2、毕业学校名称：_____。

3、学校所在地：_____省（直辖市、自治区）_____市（县）。

4、学校性质：

（1）公办高校　　（2）国有民办二级学院　　（3）民办高校

5、在各种招生类型中您属于：

（1）国家计划统招生　（2）计划外自费生　　（3）定向、委培生

（4）自学助考招生　　（3）学历文凭考试试点招生　（6）其他_____

6、您的学历是：　（1）专科生　（2）本科生　（3）硕士生　（4）博士生

7、您所学专业的名称：_____。

8、您的出生年份：19_____年；

9、您的性别和民族：（1）性别：　a)男　　b)女；

（2）民族：　a)汉族　b)少数民族；

10、您入学时的来源：

（1）来自　　a)大中城市　　b)县城　　c)乡镇　　d)农村

（2）来自_____省（自治区、直辖市）_____市（区、县）。

11、您参加高考的年份为 _____年；

您的高考成绩为_____分→　　该成绩为：　a）原始分；　b）标准分

12、您父母目前的工作状况：　父亲_____，母亲_____

（1）行政管理人员（处级或县乡科级以上干部）：　　（2）各类经理人员

（3）机关、企业、事业单位办事人员　　（4）专业技术人员
（5）个体工商人员　　（6）商业服务人员
（7）私营企业主　　（8）产业工人
（9）离退休、无业、失业、半失业　　（10）农（林、牧、渔）民

13、您父母的最高学历：　父亲_____，母亲_____
(1)研究生　(2)大学本科　(3)专科
(4)高中或中专　(5)初中　(6)小学　（7）文盲或半文盲

14、您现在的就业状况是：
（1）已经签约　　（2）已确定单位，等待签约
（3）准备从事自由职业或自主创业　　（4）等待接受单位的最后答复
（5）尚未找到接收单位　　（6）虽有愿意接受的单位，但自己不想去
（7）准备升学或出国　　（8）申请不就业
（9）其他（请说明）_____

第二部分：学业状况

15、您在学习期间担任学生干部的情况（可多选）
（1）学校级干部　（2）院、系级干部　（3）班级干部　（4）没有担任过

16、您是否为中共党员：
（1）否　（2）是→　入党时间为:19_____年

17、您的英语考级情况：
（1）大学英语 4 级　（2）大学英语 6 级　（3）没有考级　（4）其他_____

18、您是否获得过英语之外的其他等级证书或资格证书？
（1）否　（2）是→　具体名称为_____

19、您所学专业与您的兴趣的吻合状况：
（1）非常吻合　（2）基本吻合　（3）不太吻合　（4）相去甚远　（5）不知道

20、您对所在学校的教育、教学水平、办学条件的总体评价是：
（1）非常满意　（2）基本满意　（3）一般　（4）不太满意　（5）很不满意

21、关于转专业，您的情况是：
（1）从来没有过这方面想法　　（2）有过这种想法，但没有行动
（3）申请过，但没有被批准　　（4）已经有过转专业的经历

22、您是否辅修过其他专业或第二学位？
（1）否　（2）是→　辅修或第二学位专业名称是___

23、您的学习成绩在班里属于：
（1）前 25%　（2）中上 25%　（3）中下 25%　（4）后 25%

24、您的工作经历（可选多项）：
（1）入学前曾经工作过→　共_____年
（2）课余兼职半年以上→　与所学专业：　a）关系密切　b)有联

系　c)无关
（3）课余兼职半年以下——→　与所学专业：　a）关系密切　b)有联系　c)无关
（4）做过家教等临时工作
（5）参加了学校组织的实习
（6）没有工作经历
25、在本阶段（如本科四年）学习期间，您是否获得过各种奖、助学金：
（1）没有　（2）有——→　整个学习期间（如本科为四年）合计为_____元。
26、在本学习阶段（如本科四年），家庭和亲友给您的无需偿还的资助合计约_____元。
27、您是否申请过各类助学贷款：
（1）没有　（2）有——→　本学习阶段（如本科四年）合计为_____元。
28、请估计您在本阶段学习期间（如本科为四年）的各种花费状况：
（1）学费几年合计_____元　（2）辅修及课外学习班费用_____元
（3）住宿费几年合计_____元　（4）伙食费几年合计_____元
（5）其他各种费用（如交通、娱乐、通讯、文具等）几年合计约_____元。

第三部分：已有的求职经历
29、在择业过程中，您一共向_____个单位求过职，其中曾表示愿意接受您的单位为：
（1）没有　（2）1—2个　（3）3—4个　（4）5个或更多
30、您主要是通过下列哪种途径求职的（双选）：第一位_____；第二位_____
（1）学校举办的供需见面会；　（2）主管部门举办的供需见面会；
（3）有关部门的毕业分配安排；　（4）学校方面的推荐；
（5）本人的毛遂自荐、自我推销；　（6）亲友的介绍；
（7）社会上的人才招聘会；　（8）其他_____。
31、求职过程中以下哪类信息对您的帮助较大（双选）：　最重要____，其次_____。
(1)学校发布的需求信息；　(2)专门性的人才招聘信息刊物；
(3)新闻媒介的零散招聘广告；　(4)从企业得到的书面招聘广告；
(5)学校就业指导机构提供的信息　(6)在人才洽谈会获得的信息；
(7)从职业介绍机构获得的信息；　(8)亲戚、朋友或熟人介绍的信息；
(9)与用人单位的直接谈话中获得的信息；　(10)其他_____。
32、您觉得目前各方面的就业信息能否满足择业需求？
（1）完全能满足；　(2) 基本能满足；　(3) 不能满足
33、您认为下列因素对您的就业结果的影响程度如何？（在各题相应栏内画√）

	影响很大	影响较大	有一定影响	没有影响
学习成绩				
工作能力				
工作经历				
性别状况				
所学专业				
学历层次				
应聘技巧				
党员、干部				
学校名气、地位				
往届毕业生的声誉				
学校或教师的评价				
家庭背景				
社会关系				
送礼买人情				
就业信息和机会				
户口和用人指标				
“非典”疫情				

34、在您选择工作单位（对没有确定工作者而言）或确定目前工作单位（对已经确定工作单位者而言）时，下列因素的重要程度如何？（在相应栏内画√）。

	非常重要	比较重视	不太重要	不重要
工作单位的地理位置				
工作单位的所有制性质				
工作单位的规模				
工作单位的声誉				
工作稳定				
工作自由				
经济收入、福利待遇				
发展前景				
符合自己兴趣爱好				
利于施展个人的才干				
工作舒适、劳动强度低				
可兼顾亲友关系				
能获得权力和社会资源				
其他（请说明）_____				

35、到目前为止，您为求职而花费的相关费用估计共约为_____元，其中：

(1)求职简历的制作_____元　　(2)交通费　_____元

(3)招聘会门票等_____元　　(4)通讯费用　_____元

(5)人情、礼品费用_____元　　(6)其他有关费用　_____元

36、您是否参加过学校开设的就业指导课和讲座？
（1）否　　（2）是→　　共_____ 次

第四部分：工作状况（已经签约和基本确定工作去向者回答，其他同学跳到第五部分）

37、您对已确定的工作单位的了解程度如何？
（1）非常了解　（2）大致了解　（3）不太了解　（4）很不了解

38、您已经确定的就业单位在哪个地区？
（1）_____省（自治区、直辖市）_____县（市、区）
（2）该单位在：a）大中城市；　b）县城；　c)乡镇　　d)农村

39、您确定工作单位的时间是在_____年_____月；

40、根据您已知的信息，在您已确定的工作单位，您的月收入大约是_____元/月；如果加上各种补贴、福利，您的年收入总计大约为_____元/年。

41、您对自己所找到的工作的满意程度：
（1）非常满意　（2）满意　（3）一般　（4）不太满意　（5）很不满意

42、按您现在的想法，您准备在目前确定的单位工作多久？
（1）一年左右　（2）三年左右　（3）五年左右　（4）更长时间　（5）看情况

43、根据现有的了解，您将从事的工作与您所学专业的相关程度如何？
(1)非常对口　(2)基本对口　(3)有一些关联　(4)毫不相关　(5)不清楚

44、如果要胜任您即将从事的工作，您估计实际上需要哪个层次的知识和能力？
(1)初中　(2)高中或中专　(3)专科　(4)本科　(5)硕士　(6)博士

45、您已确定的工作单位的性质是：
(1)国家机关　(2)国有企业　(3)学校　(4)科研单位　(5)三资企业
(6)城市集体企业　(7)乡镇企业　(8)私营企业　(9)其他 _____

46、您已经确定的工作单位（主要指与您签约的那一级单位）的规模大约为_____人。

47、您将从事的是以下哪类工作：
(1) 各类行政管理工作（包括党政机关、事业单位、群众团体行政管理工作）
(2) 各类企业管理工作（如经理、部门经理等）
（3）各类专业技术工作（如工程师、会计师、教师、医生、律师、编辑、记者等）
（4）各类技术辅助工作（如技术员、教辅人员、护士、秘书、会计、出纳、文秘等）
（5）各类服务工作（如保安、餐饮服务、销售服务、市场营销等）
(6) 各类一线生产工人
（7）各类一线农（林、渔、牧）民
（8）其他（请说明）_____

第五部分：下一步的求职打算（目前没有确定工作、正在寻找工作者填答）

48、如果有单位愿意向您提供解决户口、档案的相对稳定的工作，您能接受的最低工资是

_____元/月：

49、在求职过程中，如果遇到不解决户口、档案问题的临时性工作，您是否会接受：

（1）不会

（2）会——→ 能接受的最低工资是_____元/月。

50、对于今后的工作与您所学专业的关系，您的态度是：

（1）一定要专业对口 （2）要求专业基本对口

（2）要与专业有一定联系 （4）与专业无关也没关系

51、您目前和将来求职时对工作地区要求是：

（1）大中城市 (2)县城 (3)乡镇 （4）农村 （5）不限地点

52、您将主要向以下哪两类单位求职？ 第一位_____；第二位_____

（1）国家机关 (2)国有企业 (3)学校 (4)科研单位 (5)三资企业

(6)城市集体企业 (7)乡镇企业 (8)私营企业 (9)其他_____

53、在今后的求职过程中，您将主要找寻哪类工作？ 第一位_____；第二位_____

(1) 各类行政管理工作（包括党政机关、事业单位、群众团体行政管理工作）

(2) 各类企业管理工作（如经理、部门经理等）

（3）各类专业技术工作（如工程师、会计师、教师、医生、律师、编辑、记者等）

（4）各类技术辅助工作（如技术员、教辅人员、护士、秘书、会计、出纳、文秘等）

（5）各类服务工作（如保安、餐饮服务、销售服务、市场营销等）

(6) 各类一线生产工人

（7）各类一线农（林、渔、牧）民

（8）其他（请说明）_____

54、在求职过程中，您遇到的最大困难是什么？最需要的帮助是什么？ 谢谢合作。

Appendix II: Survey of Willingness on Higher Education Institution Graduates Occupational Choice and Employment, 2003 (English translation of the original questionnaire)

Instruction:
Please check the appropriate answer with " √ " or with brief description and numbers.

Section I: Basic Information

1. Frequently used email address:_____
2. Name of institution:_____
3. Location of institution:_____ Province (Autonomous District or Municipals under central administration)_____City (county)
4. Ownership status of the institution: (1) Public (2) People run affiliated school under the state ownership (3) People run institution
5. Type of matriculation: (1) under state plan (2) out-of-plan fee-paying (3) ear-marked by designated job placement (4) diploma assisting (5) authorized diploma assisting (6) others
6. Highest degree obtained: (1) Associate (2) Bachelor (3) Master (4) Doctor
7. Major: _____
8. Year of Birth: 19____
9. Gender and ethnicity: (1) Gender: a) Male b) Female
   (2) Ethnicity: a) Han b) Minority:
10. Origin as of enrollment:
    (1) From a) Large and middle sized city b) county c) town / township e) village
    (2) From a)_____Province (Autonomous District or Municipals under central administration)_____city(district / county)
11. What year did you take the college entrance examination_____?
    Score of college entrance examination_____ →this score is: a) unstandardized score b) standardized score
12. Current status of your parents' occupation: Father_____, Mother_____
    (1) Government official (equivalent to county director and up); (2) Manager
    (3) Government employee (4) professionals
    (5) Self-employed (6) Third industry

employee
(7) Business owner (8) Industrial worker
(9) Retired \ Unemployed \ Semi-unemployed (10) Agricultural worker

13. Your parents' highest academic attainment:
(1) Master (2) Bachelor (3) Associate degree
(4) High school or equivalent (5) Lower secondary school (6) Primary school
(7) Illiterate or semi-illiterate

14. Your current occupation status:
(1) Signed contract with employer (2) Committed
(3) Freelancer (4) Waiting for the employment's decision
(5) Still searching (6) Not committed yet
(7) Graduate study \ overseas study (8) Not apply at all

Section II: Academic Information

15. Even been elected student carder, and what kind of carder did you do during school years?
(1) School level carder (2) Grade level carder (3) Class level carder (4) Never

16. Are you a Chinese Communist Party:
(1) No (2) Yes→Member since_____

17. College English Test (CET):
(1) Band 4 (2) Band 6 (3) Never take test (4) Others

18. Do you have other certificates than CET?
(1) No (2) Yes→Name of the certificate_____

19. How do your major coincide with your interest:
(1) Very (2) Fairly (3) Barely (4) Not at all (5) NO idea

20. How do you evaluate the teaching, facility, and quality of your institution?
(1) Very satisfied (2) Satisfied (3) Unsatisfied (5) Very unsatisfied

21. Do you even transfer to other programs?
(1) Never thing about it (2) Considered but no action
(3) Applied but not admitted (4) Transferred

22. Do you have a minor or double major?
(1) No (2) Yes→Name of the

program_____

23. Your overall academic ranking in your class:

(1) 100~75 percentile (2) 75~50% (3) 50~25% (4) 25~0%

24. Working experience (multiple):

(1) Have full time job for _____years before enter college

(2) Have part time job for more than 6 months→ and the job is: a) closely related b) related c) irrelevant→to your major

(3) Have part time job for less than 6 months→ and the job is: a) closely related b) related c) irrelevant→to your major

(4) Temporary job (e.g., Tutor)

(5) School internship

(6) None

25. Do you ever get any merit based or needed based scholarship in college?

(1) No (2) Yes→ Total amount (4 years) _____

26. Total amount of financial assistance from family and relatives (4 years) _____

27. Do you ever apply for any student loan?

(1) No (2) Yes→Total amount (4 years)_____

28. Estimate of total spending for college education(4 years):

(1) Tuition _____ (2) Tutor / minors_____

(3) Lodging expenditure_____ (4) Food expenditure_____

(5) Miscellaneous cost (e.g., transportation, recreation, telephone, school supply ant etc.)_____

Section III: Job Seeking Experience

29. I applied for ____ firms and organizations. How many of these firms and organizations respond positively?

(1) None (2) 1~2 (3) 3~4 (4) 5 and up

30. Means of job seeking (bi-choice): _____and_____

(1) School job fair (2) Official job fair

(3) Though official job assigning (4) School recommendation

(5) Self promotion (6) Family and social connections

(7) Commercial job fair (8) Others_____

31. Which two kinds of information are most helpful to your job searching: 1st____and 2nd_____

(1) Job information disseminated by school (2) Career development magazines

(3) Commercial classified ads (4) Direct mailed want ads
(5) Information from school career center (6) Job fair
(7) Head hunting agency (8) Family or social connections
(9) Direct contact the employer (10) others_____

32. Are you satisfied with the information provided by various channels?
(1) Very (2) Fairly (3) Not at all

33. How do you evaluate the following elements' impact to your career? (Check √ in the cell)

	Decisive	Influential	Somewhat	Not at all
Academic performance				
Working ability				
Working experience				
Gender				
Major / Program				
Degree				
Job searching skills				
CCP membership / card				
Reputation of school				
Reputation of former graduates				
School/teacher recommendation				
Socioeconomic status				
Social resources				
P.R. relationship				
Career information				
Quota / residency registration				
Outbreak of SARS				

34. If you do not commit to any employer, how do you weight the importance of the following elements of your future employer? (Check √ in the cells)

	Very	Fairly	Barely	Not at all
Location				
Ownership status of the employer				
Number of employee				
Reputation of the emploer				
Job security				
Commitment to the employer				
Salary and benefits				
Future personal development				

Consistency with personal interest				
Room of self-realization				
Working environment / work load				
Family care taking				
Social power and resources				
Others_____				

35. To date, the rough figure of your job seeking cost_____, and the breakdown is:

(1) Resume_____ (2) Transportation_____

(3) Ticket to job fairs_____ (4) Communication cost _____

(5) P.R. relations _____ (5) others _____

36. Did you attend career seminar offered by your school?

(1) No (2) Yes→_____times

Section IV: Information of Job (Please skip to section V, if you don't have a commitment yet.)

37. How do you know about your employer?

(1) Very much (2) In general (3) Barely (4) Not quite

38. Locality of your employer:

(1) _____Province (Autonomous District or Municipals under central administration)_____county (City / District)

(2) Located in a) Large and middle sized city b) county c) town / township e) village

39. When did you get the confirmation of your job? _____,_____(year, month)

40. You estimated salary / month_____; plus benefit and other package, your annual income will be _____

41. How are you satisfied with your job?

(1) Very satisfied (2) satisfied (3) acceptable (4) Barely (5) Not at all

42. How long will you plan to work for the current employer?

(1) 1 year (2) 3 years (3) 5 years (4) more than 5 years (5) Not sure yet

43. How does your current occupation match to your major?

(1) Very (2) Basically (3) Somewhat (4) Not at all (5) Not sure yet

44. What kind of academic attainment do you think your occupation requires?

(1) Lower-secondary (2) High school (3) Associate degree (4) Bachelor (5) Master

(6) Doctor

45. Ownership status of your employer:
(1) Government (2) State owned (3) Public school (4) Public research institute (5) Foreign enterprise (6) Private cooperate (7) Public cooperate (8) Private firm (9) other_____
46. Total number of employee of your employer:_____
47. You job type:
(1) Administrative (government, public institution, NGOs and etc.)
(2) Managerial (manager, director)
(3) Professional (engineer, accountant, teacher, doctor, lawyer, journalist and etc.)
(4) Professional assistant (technician, nurse, secretary, treasurer and etc.)
(5) Third industry (security, salesperson, and etc.)
(6) Industrial worker
(7) others_____

Section V: Job Seeking Plan (for those who have not yet found a job)

48. If the employer will offer you a package including residency registration (hu kou) settlement, relative stable salary, what is the minimum wage can you accept?_____/month
49. Will you accept a temporary job offer without residency registration (hu kou) settlement?
(1) No (2) Yes→with_____/month minimum wage
50. How do you think of your future job should match your major?
(1) Must match (2) Generally match (3) Somewhat related (4) Doesn't matter
51. Ideal locality of future employment:
(1) Large and middle sized city (2) County (3) Town/township (4) Village (5) Doesn't matter
52. What are the major types of employers you are looking for?_____1st _____2nd
(1) Government (2) State own enterprise (3) School (4) Research institute (5) Foreign enterprise (6) County-level public corporate (7) Township public corporate (8) Private enterprise (9) others_____
53. What type of job will you be looking for in the future? _____1st _____2nd
(1) Administrative (government, public institution, NGOs and etc.)

(2) Managerial (manager, director)
(3) Professional (engineer, accountant, teacher, doctor, lawyer, journalist and etc.)
(4) Professional assistant (technician, nurse, secretary, treasurer and etc.)
(5) Third industry (security, salesperson, and etc.)
(6) Industrial worker
(7) Farmer / Fisherman
(8) others_____

54. What is the most difficult thing do you encounter in your job searching and, what kind of assistant is most needed?

Thank you for your cooperation!