Human Capital in China

China Center for Human Capital and Labor Market Research
Central University of Finance and Economics
Beijing, China

October 6，2009

Prepared for
International Symposium on
Measuring Human Capital and Its Impact on Welfare and Growth:
Inaugurating the China Human Capital Index
October 9-10, 2009
Beijing, China

Disclaimer

This report is drafted
with excellent assistance from other project team members
by

Haizheng Li

Director and Professor

China Center for Human Capital and Labor Market Research (CHLR)

&

Georgia Institute of Technology

Barbara Fraumeni

Senior Fellow, CHLR

&

University of Southern Maine

Zhiqiang Liu

Special-term Professor, CHLR

&

State University of New York at Buffalo

Xiaojun Wang

Special-term Professor, CHLR

&

University of Hawaii at Manoa

This project is funded by

National Natural Science Foundation of China

and

Central University of Finance and Economics

Brief Introduction of
China Center for Human Capital and Labor Market Research

Established in March 2008, China Center for Human Capital and Labor Market Research (CHLR) at the Central University of Finance and Economics is an essential part of the Advantageous Program Platform in Economic and Public Policy. It is an international research center for the study of human resources, labor markets, and their impact on economic development, focusing on China and related economies. The CHLR has master's program, doctoral program, and post-doctoral program. Our advisory board includes two Nobel laureates, among many other internationally renowned scholars.

Our major research areas include: human capital investment; human capital mobility; human capital innovation and technology; and health and human capital.

Faculty members and research fellows of the CHLR are mostly economics doctorates from the U.S., many are tenured professors at U.S. universities. The CHLR Special-term Director, Dr. Haizheng Li, is also Associate Professor at the School of Economics, Georgia Institute of Technology. Currently the Center has 3 full-time faculty, 4 special-term faculty, 7 senior research fellows, and 3 research fellows.

The Center has established a graduate program following the international standard. The curriculum and instruction are rigorously designed, and courses are usually taught by a team of international scholars. All the courses are taught in English. Currently the CHLR has 25 master students, 2 doctoral students, and 1 post-doctoral fellow.

The Center is foremost an international research institution. We not only have a team of international scholars and train graduate students by international standard, we also introduce international management into Center's daily business. Faculty researches are evaluated by a standard adopted by similar institutions in the U.S., and all graduate students are involved in research projects soon after they join the program.

Since the Center was established about one and a half years ago, it has received two research grants from the China National Natural Science Foundation. CHLR Director, Professor Haizheng Li, is the principle investigator of both projects. The first project is entitled "China Human Capital Measurement and Human Capital Index Project". This project involves all full-time faculty, special-term professors, graduate students, and staff members. The second project is supported by the Director Fund of the China National Natural Science Foundation, entitled "The Feasibility Analysis of Supporting Economic and Managerial Data Base". This is a joint project with Qinghua University School of Economy and Management and Renmin University Business School.

Introduction to
China Human Capital Index Project

"China Human Capital Measurement and Human Capital Index Project" is funded by China National Natural Science Foundation and Central University of Finance and Economics, conducted by China Center for Human Capital and Labor Market Research (CHLR). The goal of this project is to establish China's first set of systematic and scientific measurements of human capital and quantify its distribution and dynamics. The Indexes, once established, can be used to support empirical research as well as government policy-making. In addition, the China human capital index we are constructing is aimed at becoming an important part of the nascent international human capital measurement system, and eventually being incorporated into the National Income Accounting system.

This project is led by CHLR Director, Professor Haizheng Li. Professor Barbara Fraumeni, who did the pioneer work in developing the popular Jorgenson-Fraumeni method of calculating human capital stock, and all faculty members and graduate students at the CHLR participated in the project.

This project requires a huge amount of data collection and processing. After one year of daily effort, we have obtained China's total human capital stock series from 1985 to 2007. We have also calculated disaggregated values by location (i.e. urban and rural) and gender, and projected the series until 2020. Our results have seen rising attention from international organizations such as the OECD, and we are actively looking for opportunities of more international collaboration.

Research Team Members

Principle Investigator

Haizheng Li	Special-term Director, CHLR at CUFE Associate Professor of Economics, Georgia Institute of Technology

Members

Professors and Staff

Ake Blomqvist	Professor, CHLR
Belton Fleisher	Special-term Professor and Senior Fellow, CHLR Professor of Economics, Ohio State University
Barbara Fraumeni	Senior Fellow, CHLR Associate Dean and Professor of Public Policy, Muskie School of Public Service, University of Southern Maine
Zhiqiang Liu	Special-term Professor, CHLR Associate Professor of Economics, State University of New York at Buffalo
Xiaojun Wang	Special-term Professor, CHLR Associate Professor of Economics, University of Hawaii at Manoa
Kang-Hung Chang	Assistant Professor, CHLR
Song Gao	Assistant Professor, China Academy of Public Finance and Public Policy, CUFE
Zhiyong Liu	Instructor, Hunan University of Commerce
Ruiju Wang	Executive Assistant to Director, CHLR
Hao Deng	Graduate Coordinator, CHLR

Graduate Students

CHLR	Yunling Liang (Ph.D.), Huajuan Chen, Yuhua Dong, Mengxin Du, Jinquan Gong, Jingjing Jiang, Rui Jiang, Qian Li, Sen Li, Chen Qiu, Xinping Tian, Mo Yang
Georgia Institute of Technology	Yuxi Xiao

Human Capital in China

China Center for Human Capital and Labor Market Research
Central University of Finance and Economics

Contents

Executive Summary

In this project we estimate China's human capital stock from 1985 to 2007 based on the Jorgenson-Fraumeni lifetime income approach. An individual's human capital stock is equal to the discounted present value of all future incomes he or she can generate. In our model, human capital accumulates through formal education as well as on-the-job training. The value of human capital is assumed to be zero upon reaching the mandatory retirement ages.

China's total real human capital increased from 26.98 billion yuan in 1985 (i.e., the base year) to 118.75 billion yuan in 2007, implying an average annual growth rate of 6.78%. The annual growth rate increased from 5.11% during 1985-1994 to 7.86% during 1995-2007. Per capita real human capital increased from 28,044 yuan in 1985 to 106,462 yuan in 2007, implying an average annual growth rate of 6.25%. The annual growth rate also increased from 3.9% during 1985-1994 to 7.5% during 1995-2007. Therefore, although population growth contributed significantly to the total human capital accumulation before 1994, per capita human capital growth was primary driving force after 1995. The substantial increase in educational attainment during 1985-2007 contributed significantly to the growth in total and per capita real human capital.

Since human capital accumulation was slower than GDP growth and physical capital accumulation, the ratio of human capital to GDP fell from 30 in 1985 to 18 in 2007, the ratio of human capital to physical capital declined from 16 in 1985 to 11 in 2007. These values are not far away from those obtained in studies on other countries. An important unanswered question is whether optimal values of human capital relative to physical capital and GDP can be defined in relationship to sustainable economic growth.

In 2007, total male human capital was about twice that of total female human capital, this gap is slightly larger than in 1985. However, female per capita human capital is nearly 72% of male per capita human capital in 2007, indicating that most of the gap in total human capital can be attributed to differences in population, returns to schooling and work experience, and mandatory retirement age. Rural total human capital was greater than that of urban in 1985, but urban overtook rural in the early 1990s, and by 2007 urban total was about twice of rural total. Urban per capita human capital increased from 47,874 yuan in 1985 to 154,803 yuan in 2007, while rural per capita human capital increased from 21,856 yuan to 66,164 yuan. The rural-urban gap increased by about 3 percentage points (i.e., the rural-urban per capita human capital ratio was 45.7% in 1985 and 42.7% in 2007).

In our projection from 2007 to 2020, total human capital will grow at a much slower annual rate of 0.61%. This is mainly because we assume future parameters and values will remain the same as their 2007 values. Urban total human capital will continue to rise, while rural total human capital will slowly decline, mainly due to continued migration and urbanization. Per capita human capita, however, will remain constant in the rural area and will grow slowly in the urban area.

I. Introduction

Since the concept of human capital was introduced to modern economic analysis by Schultz (1961) and Becker (1964), it has been widely used in academic studies and policy analysis. Human capital is probably "the most important and most original development in the economics of education" in the second part of the 20th century (Coleman, 1990, page 304). The latest definition of human capital from the Organization for Economic Co-operation and Development (OECD) is "The knowledge, skills, competencies and attributes embodied in individuals that facilitate the creation of personal, social and economic well-being" (OECD, 2001, page 18). In most countries, human capital accounts for more than 60% of the nation's wealth, which includes natural resources, physical capital and human capital (World Bank, 1997).

It is generally believed that human capital is an important source of economic growth and innovation, an important factor for sustainable development, and for reducing poverty and inequality (see, for example, Stroombergen et al., 2002, and Keeley, 2007). For example, the detailed analysis of human capital accounts for Canada, New Zealand, Norway, Sweden, and the United States unanimously shows that human capital is a leading source of economic growth.[1]

In China, since the start of economic reforms, the economy has grown at a dramatic rate. It is believed that human capital has played a significant role in the Chinese economic miracle (see, for example, Fleisher and Chen, 1997, and Démurger, 2001). Additionally, studies show that human capital also has an important effect on productivity

---

[1] These include Jorgenson-Fraumeni (J-F) accounts for Canada (Gu and Ambrose 2008), New Zealand (Le, Gibson, and Oxley 2005), Norway (Greaker and Liu 2008), Sweden (Alroth 1997), and the United States (Jorgenson and Fraumeni 1989, 1992a, 1992b, and Christian 2009).

growth and on reducing regional inequality in China (Fleisher, Li and Zhao, 2009).

Despite the important role of human capital in the Chinese economy, however, until now, there has been almost no comprehensive measurement of the total stock of human capital in China. Human capital measures for China are central to any understanding of the global importance of human capital for a number of reasons. First, China is the most populous country in the world. It is important to understand the dynamics of human capital caused by demographic changes (for example, due to one-child policy, migration, and urbanization) and by the rapid expansion of education during the course of economic development. Second, such measures would allow for better assessment of the contribution of human capital to growth, development, and social well-being in empirical and theoretical research. Construction of human capital measures is an important step in assessing the contribution of human capital to economic growth. Currently, only partial measurement of human capital, such as education characteristics, has been used in such studies.

Additional benefits from human capital measures include the provision of useful information for policy makers, such as assessing how education policies of central and local governments affect the accumulation of human capital. This is especially important, given the long-term nature of human capital investment. For example, since the early 1980s, there has been a remarkable increase in the educational attainment of the Chinese population. In 1982 the largest population mass was concentrated in the “no schooling” category (Figure III.1.4). By 2007 the largest population mass was concentrated in the “junior middle” school category (Figure III.1.7). Developing comprehensive measures of human capital in China provides the necessary early work for constructing China’s human capital account and for eventually incorporating human capital into the national accounting so that China can join the international OECD initiative. It would facilitate international comparison of human capital accumulation and growth across nations.

There is an ongoing international effort in developed countries to measure a nation's total human capital stock and to develop national human capital accounts. For example, the United States formed the Committee on National Statistics' Panel to Study the Design of Nonmarket Accounts (Abraham 2005, and Christian 2009); in early 2008, Statistics Canada set up a program "Human Development and its Contribution to the Wealth Accounts in Canada" (Gu and Wong 2008); Australian Bureau of Statistics (Wei 2008), Statistics Norway (Greaker and Liu 2008) and New Zealand (Le, Gibson, and Oxley 2005), have also established similar research program on the measurement of human capital. In addition, seventeen countries: Australia, Canada, Denmark, France, Italy, Japan, Korea, Mexico, Netherlands, Norway, New Zealand, Poland, Spain, the United Kingdom, the United States, Romania, and Russia, and two international organizations Eurostat and the International Labour Organization, have agreed to join the OECD consortium to develop human capital accounts. A researcher from Statistics Norway, Gang Liu, is at the OECD as of October 1, 2009 for nine months to coordinate this effort. The work of this consortium will facilitate cross-country comparisons. In addition, the Lisbon Council European Human Capital Index has been constructed for the 13 European Union (EU) states and 12 Central and Eastern European states (See Ederer 2006 and Ederer *et. al.* 2007). Developed countries have obviously realized the importance of monitoring human capital accumulation, while most developing countries have yet to start such projects, including China.

Until now, there has been no systematic effort to construct comprehensive measures of the total human capital stock in China. There are a few studies on human capital measurement published in Chinese journals. For example, Zhang (2000) and Qian and Liu (2004) calculated China's human capital stock based on total investment (cost-side); others, such as Zhu and Xu (2007), Wang and Xiang (2006), estimated human capital from the income side. Zhou (2005) and Yue (2008) used some weighted average of human capital attributes to construct a measurement. In most cases, these studies partially measure human capital based on some education characteristics such as average education, for example, Cai (1999), Hu (2002), Zhou (2004), Hou (2000), Hu (2005), etc.

While the above studies did contribute to the understanding of human capital in China, there are major limitations. First, there has been no comprehensive and systematic measurement of the total human capital stock in China from the 1980s up to date, especially on the changes of human capital in rural and urban areas and for males and females respectively. Second, the methodology used has been limited by data availability, feasibility of parameter estimation, and some technical treatment difficulties. Thus, there has no exact implementation of internationally recognized methods to China's data for human capital estimation.

We attempt to construct a comprehensive measurement of human capital in China by applying the methods used in other countries after modifying them to fit China's special cases. We estimate total human capital at the national level, for male and female, for urban and rural areas from 1985 to 2007. Our estimates include nominal values, real values, indexes, and quantity measures. We mostly adopted the Jorgensen-Fraumeni (J-F) lifetime income based approach, which has been widely used in other countries.

In addition to a full-implementation of the J-F approach to China's data to estimate the human capital series, another contribution of this study is that we combine micro-level survey data in human capital estimation to mitigate the lack of earnings data in China. In particular, we apply the

Mincer equation to estimate earnings by using various available household survey data. Thus, it is possible to integrate the changes of returns to education and experience (on-the-job-training) into our estimates during the course of economic transition.

Moreover, by separating the calculation of human capital for urban and rural areas, we are able to capture the changes caused by rapid urbanization as well as by the large scale rural-urban migration since the start of economic reform in China. This framework is not only important for any transitional economy because of its changing economic structure and migration, it can also at least partially measure the effect of another type of human capital investment—migration, which helps realize higher value of one's human capital.

The rest of this report is arranged as follows. Section II discusses methodology for human capital measurement. Section III describes our data and data treatments. The estimated results of human capital are reported in Section IV. Section V concludes. All data and technical details are reported in the appendixes.

II. Methodology

In general, human capital can be produced by education and training (child bearing and rearing are investments that increase future human capital), as well as by job turnover and migration that help to realize the potential value of human capital. Like physical capital stock, the human capital can be valued using two methods: i) it can be valued as the sum of investment, minus depreciation, added over time to the initial stock; ii) it can be valued as the net present value of the income flow it will be able to produce over an assumed lifetime. The first method, the perpetual inventory method, is used in the cost approach; while the second method is the income-based approach (this method is used to estimate the value of most natural resources). When human capital is measured using the perpetual inventory approach, only costs or expenditures are included in investment. When physical capital is measured, investments are valued at their purchase price which is not generally available for human capital.

There are several measures of human capital commonly adopted by researchers:

(1) The lifetime income approach of Jorgenson and Fraumeni (1989, 1992a, 1992b);
(2) The cost approach of Kendrick (1976);
(3) The indicator approach;
(4) Laroche and Merette (2000) construct indexes with either relative wage weights or relative lifetime income weights;
(5) The Lisbon Council's approach (2006) is described as an example of the indicator approach;
(6) The World Bank residual approach (2006).

The approach of Jorgenson-Fraumeni is discussed further next.

II.1 Jorgenson-Fraumeni income-based approach

The Jorgenson and Fraumeni (J-F) income-based approach is the most widely used method in estimating human capital stock, and has been adopted by a number of countries in constructing human capital accounts (see footnote 1 for examples). The advantages of this approach are that it has a sound theoretical foundation and that the data and parameters are relatively easier to obtain than they are for other approaches.

When estimating lifetime income to calculate human capital, an important issue is that income (or implicit income) can be generated from both market and non-market activities. Market activities of individuals produce goods and services, foster innovation and growth through managerial and creative activities, and generate income that allows for the acquisition of market goods and services. Nonmarket activities of individuals include household production, e.g., cooking, cleaning, and care-giving. Investment is generated from both market and nonmarket activities. Because household production activities are difficult to quantify and value and require time-use estimates, we have opted to exclude them in this first approximation to estimating China's human capital.[2] The J-F approach imputes expected future lifetime incomes based on survival, enrollment, and employment probabilities. Expected future wages and incomes are estimated from the currently observed wages and incomes of the cross section of individuals who are older than a given cohort at the time of observation. Future incomes are augmented with a projected labor income growth rate and discounted to the present with

[2] Among the most recent human capital estimates, i.e., Gu and Ambrose (2008), Greaker and Liu (2008) and Christian (2009), only Christian, for the United States, includes a full set of nonmarket activities and estimates human capital for those too young to go to school or to perform market work.

a constant interest rate. Estimation is conducted in a backward recursive fashion, from those aged 75, 74, 73, and so forth to those aged 0.[3]

With the J-F income-based approach, we first need data or estimates of individual's annual market labor income per capita. Then lifetime incomes are calculated by a backward recursion, starting from the oldest cohorts in the population. The life cycle is divided into five stages, and the equations used for calculating the lifetime expected incomes are as follows.

The first stage is no school and no work:

$$mi_{s,a,e} = sr_{s,a+1} \times mi_{s,a+1,e} \times \frac{real \quad income \quad growth \quad rate}{discount \quad rate}$$

where the subscripts *s, a,* and *e* denote sex, age and educational attainment respectively. *mi* stands for lifetime market labor income per capita, and *sr* is the survival rate, defined as the probability of becoming a year older.

The second stage is school but no work:

$$mi_{y,s,a,e} = [senr_{y+1,s,a+1,e+1} \times sr_{y+1,s,a+1} \times mi_{y,s,a+1,e+1} + \left(1 - senr_{y+1,s,a+1,e+1}\right) \times sr_{y+1,s,a+1} \times mi_{y,s,a+1,e}] \times \frac{real \quad income \quad growth \quad rate}{discount \quad rate}$$

where *senr* is school enrollment rate and subscript *enr* refers the grade level of enrollment, the probability that an individual with educational attainment *e* is enrolled in education level *e*+1.

The third stage is school and work. With *ymi* denoting annual market income per capita, the equation can be written as:

$$mi_{y,s,a,e} = ymi_{y,s,a,e} + [senr_{y+1,s,a+1,e+1} \times sr_{y+1,s,a+1} \times mi_{y,s,a+1,e+1} + \left(1 - senr_{y+1,s,a+1,e+1}\right) \times sr_{y+1,s,a+1} \times mi_{y,s,a+1,e}] \times \frac{real \quad income \quad growth \quad rate}{discount \quad rate}$$

[3] The J-F inclusion of nonmarket lifetime income and expected lifetime income for youngsters produces human capital estimates that are notably higher than those in the studies mentioned above who have adopted the J-F methodology.

The fourth stage is work but no school:

$$mi_{s,a,e} = ymi_{s,a,e} + sr_{s,a+1} \times mi_{s,a+1,e} \times \frac{real \quad income \quad growth \quad rate}{discount \quad rate}$$

The fifth and final stage is retirement or no school or work:

$$mi_{s,a,e} = 0$$

Similar equations can be applied to estimate lifetime nonmarket labor income, which can be added to lifetime market labor income to give total lifetime labor income.

To depict the growth rate of human capital, quantity indexes are introduced by J-F approach. Two kinds of quantity indexes are estimated for China.

(1) Gender-based quantity index

In this case, two weighted growth rates are used to create the Divisia index according to the formula:

$$Migrowth_y = \frac{1}{2}\sum_s \left(Mishare_{y,s} + Mishare_{y-1,s}\right) \times \left[\ln\left(Pop_{y,s}\right) - \ln\left(Pop_{y-1,s}\right)\right]$$

where s=male or female, y denotes year, $Migrowth_y$ is the growth rate in year y, $Mishare_{y,s}$ is the share of lifetime income for males or females in year y (or y-1 when that subscript is used). $Pop_{y,s}$ is the number of males or females in year y (or y-1 when that subscript is used).

(2) Education level-based quantity index

In this case, five weighted growth rates in all years or six weighted growth rates after 2000 are used to create the Divisia index. The formula is:

$$Migrowth_y = \frac{1}{2}\sum_e \left(Mishare_{y,e} + Mishare_{y-1,e}\right) \times \left[\ln\left(Pop_{y,e}\right) - \ln\left(Pop_{y-1,e}\right)\right]$$

where e denotes education levels, including primary school, junior middle school, senior middle school, etc. The other notation is the same as before.

II.2 Cost approach

Kendrick is an early pioneer in the construction of human capital accounts. Kendrick (1976) estimates both tangible and intangible human capital. Tangible human capital includes child rearing costs. Intangible human capital includes education, training, medical, health and safety expenditures, and mobility costs. Human capital stocks are created using a perpetual inventory method where investment expenditures are cumulated and existing stocks are depreciated. Implementation of a Kendrick approach for China is difficult as Kendrick's human capital investment is the sum of a long list of human capital related costs, and reliable data on such information is only available for the most recent decades.

Tangible human capital investment is average lifetime rearing costs including expenditures on food, shelter, health, schooling, and so on. The cost of parental time is not included in this measure. Intangible human capital investment in formal and informal education includes both private and government costs. Private formal education costs include net rental for private education sector's plant and equipment and students' expenditures on supplies. The estimate for the cost of rentals of books and equipment depends on a student's imputed potential compensation. Government formal education costs include all types of expenditure, including those for construction. Personal informal education expenditures include a portion of those for radio, TV, records, books, periodicals, libraries, museums, and so forth. Business and institutional expenditures include a portion of those for media expenditures. Religious education expenditures are imputed from figures on religious class attendance and imputed interest on plant and equipment of religious organizations. Government expenditures include those for library, recreation costs and military expenditures.

Intangible human capital investment in training values initial nonproductive time and nonwage costs and includes explicit training expenditures. Both specific and general training is captured, as well as military training. A substantial fraction of medical, health and safety expenditures, which are split between investment and preventive expenditures, are by governments. Annual rental costs for plant and equipment are imputed when not available.

Kendrick considers his human capital mobility investment estimates to be tentative. These include unemployment, job-search, hiring, and moving costs, for both residents and immigrants. Depreciation is estimated using the depreciation methodology most widely used at the time of his research: A double declining balance formula with a switch to a straight-line method. Lifetimes in these formulas are assumed to be the reciprocal of the percentage of persons in the group.

Kendrick nominal human capital is about five times Gross Domestic Product. However, Jorgenson-Fraumeni human capital is substantially larger than Kendrick human capital.[4] The Kendrick approach covers detailed aspects of human capital formation from the cost side and provides a very complete menu for sum up all related cost to estimate the value of human capital. Yet, the data requirement is enormous, for example, we may need to get government statistics ninety years back to do the calculation. This is impossible, given the People's Republic of China is only 60 years old in 2009. Additionally, it lacks guideline for many technique treatments, such as for the split of health expenses between investment and preventative costs. Therefore, we do not adopt it here for our calculation.

II.3 Indicator approach

An example of an indicator approach is the Human Capital Index of the Lisbon Council. It is a human capital input cost, or cost of creation approach. This index has been constructed for the 13 European Union (EU) states and 12 Central and Eastern European states as previously noted.[5]

---

[4] See table 37 of Jorgenson-Fraumeni (1989).

[5] See Ederer (2006) and Ederer *et. al.*(2007). The 2006 paper states that the index was developed by the German think tank Deutschland Denken. In addition the paper states that the paper is part of a research project undertaken by several individuals in the think tank and with the institutional support of Zeppelin University.

The Human Capital Endowment measure is an input to two of the other three components of the overall European Human Capital Index. The Human Capital Endowment measure sums up expenditures on formal education and the opportunity cost of parental education, adult education, and learning on the job. Parental education includes teaching their children to speak, be trustful, have empathy, take responsibility, etc. The Human Capital Utilization Index is the endowment measure divided by total population and the Human Capital Productivity Measure is Gross Domestic Product (GDP) divided by the endowment employed in the country.

Finally the Demography and Employment measure estimates the number of people who will be employed in the year 2030 in each country by looking at economic, demographic, and migratory trends.[6] As it has cost components and index components, it is best viewed as a blend of a cost approach and an indicator approach. Since the technique details for this approach have not been released, we do not apply it here in our calculation.[7]

II.4 Attribute-based approach

The attribute-based approach is usually considered to be a variant of the income-based approach (Le, Gibson and Oxley 2003, 2005). However, it constructs an index value of human capital instead of a monetary value in other income-based methods. The primary advantage of an index value is that it nets out the effect of aggregate physical capital on labor income, therefore this measure captures the variation in quality and relevance of formal education across time and country.

[6] Ederer (2006), p. 4 and p. 20.
[7] We have discussed with Dr. Ederer on possible collaboration of applying the China data to their method in the future.

Based on the pioneer work of Mulligan and Sala-i-Martin (1997), Koman and Marin (1997) applied the attribute-based method to Austria and Germany. However, our method is akin to Laroche and Merette (2000) in that we also incorporate work experience into the model along with formal education. That is, we also emphasize informal channels, such as work experience, in the accumulation of human capital.

Specifically in this method, the logarithm of human capital per capita in a country at any time is computed using the following formula

$$\ln\left(\frac{H}{L}\right) = \sum_e \sum_a \omega_{e,a} \ln(\rho_{e,a})$$

$$\omega_{e,a} = \frac{e^{\sum_s \left(\beta_s e + \gamma_s Exp + \delta_s Exp^2\right)\varphi_{s,a}} L_{e,a}}{\sum_e \sum_a e^{\sum_s \left(\beta_s e + \gamma_s Exp + \delta_s Exp^2\right)\varphi_{s,a}} L_{e,a}}$$

where *e* and *a* denote years of formal schooling and age, respectively. $\rho_{e,a} = L_{e,a} / L$ is the proportion of working age individuals of age *a* with *e* years of schooling. $\omega_{e,a}$ is the efficiency parameter defined as proportion of wage income of workers of age *a* with *e* years of schooling in the total wage bill of the economy. e*xp* represents work experience, which is defined as *a*-*e*-6. *s* is a gender index and $\omega_{e,a}$ is the share of men and women of age *a* in the population. Parameters *β*, *γ* and *δ* are estimates from a standard Mincer equation. The parameter *β* is often considered to be the rate of return to one more year of formal education.

In order to implement this method, we need to construct a population data set by age, gender and educational attainment for each year we study. Secondly, we need two sets of estimates from Mincer equations for each year, one for each gender. It is feasible to calculate a human capital measure based on this approach. The major issue is that in this setup, the measurement is actually a Cobb-Douglas formula. In other words, the proportions of different education groups by construction are not "perfect substitutes." When the share of one education group increases, it could cause the total measurement to decline. For example, if we increase the proportion of population with higher education, the measurement should increase as the overall education get higher, but it could decline due to the Cobb-Douglas formulation. This happened in our calculation. Since we believe that an education-based human capital measurement should be a monotonically increasing function of the overall education, we do not report the results of the attribute-based approach. In our future work we plan to modify the structure, using, for example, average years of schooling.[8]

II.5 Residual approach

The World Bank (2006) uses a residual approach to estimating human capital for 120 countries. Due to data and methodological limitations, total wealth in the year 2000 is measured as the net present value of an assumed future consumption stream. The value of produced capital stocks is estimated with the perpetual inventory method. Produced capital includes both structures and equipment. Natural capital is valued by taking the present value of resource rents. Natural capital includes nonrenewable resources, cropland, pastureland, forested areas, and protected areas. Intangible capital is equal to total wealth minus produced and natural capital. Intangible capital is an aggregate which includes human capital, the infrastructure of the country, social capital, and the returns from net foreign financial assets. Net foreign financial assets are

[8] This point was confirmed by email communication with Dr. Reinhard Koman.

included because debt interest obligations will affect the level of consumption. Intangible capital represents greater than 50% of wealth for almost 85% of the countries studied.

Using a net present value approach to estimate total wealth requires assumptions about the time horizon and the discount rate. The World Bank chooses 25 years as the time horizon as it roughly corresponds to one generation. It chooses a social discount rate rather than a private rate as governments would use a social discount rate to allocate resources across generations. The social discount rate is set at 4%, which is at the upper range of estimates it reviewed for industrialized countries. The same rate is used for all countries to facilitate comparisons across countries.

A Cobb-Douglas specification is employed to estimate the marginal returns and contribution of three types of intangible capital in the model. The model independent variables include per capita years of schooling of the working population, human capital abroad, and governance/social capital. Human capital abroad is measured by remittances by workers outside the country. Governance/social capital is measured with a rule of law index. Although the marginal return to human capital in the aggregate is the highest of the three included intangible capital components, the contribution decomposition demonstrates that the relative contributions can differ significantly across countries (World Bank, 2006, chapter 7).

III. Data

III.1 Population

In order to implement the various methods used in estimating human capital, we first and foremost need annual population data by age, sex, and educational attainment. We construct such data sets according to the following procedure.

First, data sets are available for the years 1982, 1987, 1990, 1995, 2000, and 2005. They are reported in various issues of Population Census, Population Sampling Survey, and Population Yearbooks. The data sets also contain disaggregated numbers for urban and rural populations.

For all other years, we collect population data by age and sex from various issues of China Population Yearbooks. Then we combine birth rate (China Statistical Yearbook), mortality rate by age and sex (China Population Yearbook), and enrollment (including new enrollment and graduation, China Education Statistical Yearbook) at different levels of education to impute population by age, sex and educational attainment for each and every year. We define the following levels of educational attainment: illiterate (no schooling), primary school (Grade 1-6), junior middle school (Grade 7-9), senior middle school (Grade 10-12), and college and above. From 2000 on, additional information makes it possible to separate the population at the level of college and above into two: one is college, and the other is university and above.

Specifically, we use the following perpetual inventory formula to deduce population by age, sex and educational attainment in missing years:

$$\begin{aligned} L(y,e,a,s) = {} & L(y-1,e,a,s)\cdot\big(1-\delta(y,a,s)\big)+IF(y,e,a,s) \\ & -OF(y,e,a,s)+EX(y,e,a,s) \end{aligned}$$

$L(y,e,a,s)$ is the population in year y at education level e, with age a and sex s. $\delta(y,a,s)$ is the mortality rate in year y, with age a and sex s. $IF(y,e,a,s)$ and $OF(y,e,a,s)$ are inflow and outflow of this particular group. For example, inflow would include individuals just achieved this level of education, while outflow would include those who just achieved the next level of education. $EX(y,e,a,s)$ is a discrepancy term. Moreover,

$$IF(y,e,a,s) = \lambda(y,e,a,s) \cdot ERS(y,e,s)$$
$$OF(y,e,a,s) = \lambda(y,e+1,a,s) \cdot ERS(y,e+1,s)$$
$$\sum_a \lambda(y,e,a,s) = 1$$

ERS is the matriculation at education level e, λ is the age distribution at education level e. In order to obtain accurate estimate for λ, we use both microeconomic data sets (China Health and Nutrition Survey and China Household Income Project) and macroeconomic data sets (China Education Statistical Yearbook). Details can be found in Appendix A.

Next we discuss several salient features of China's population growth, especially the educational attainment by age, sex, and location (i.e. urban and rural). First of all, during our sample period, China's total population increased from 1.02 billion in 1982 to 1.32 billion in 2007. The urban population increased by 379 million, while the rural population decreased by 74 million (Figure III.1.1). As a result, the urban share in the total population rose from 21% in 1982 to 45% in 2007. The male and female population almost rose at the same pace, with the male's share remained at around 51% (Figure III.1.2).

Figure III.1.1 Population in China, 1982-2007

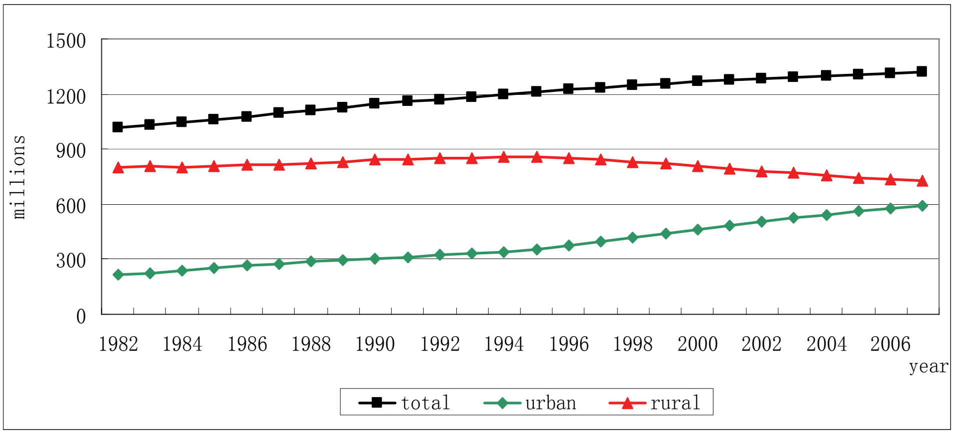


Figure III.1.2 Population in China, 1982-2007

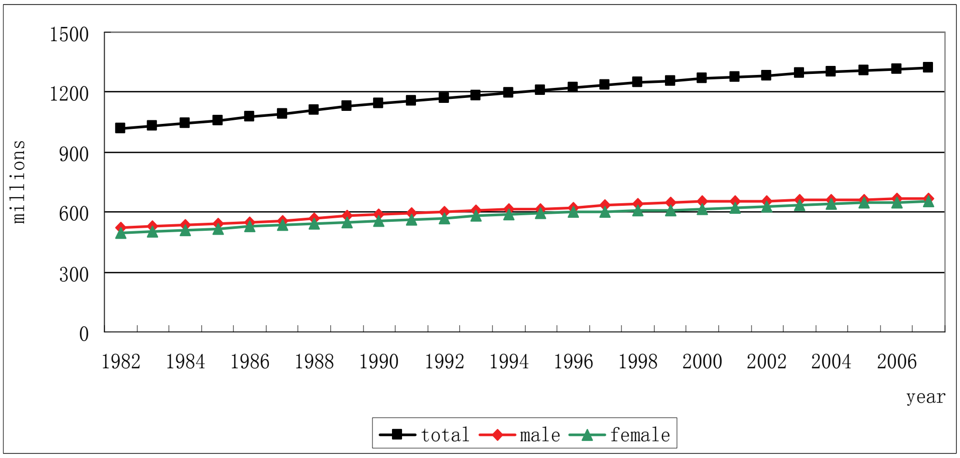


Figure III.1.3 Population by educational attainment, 1982-2007

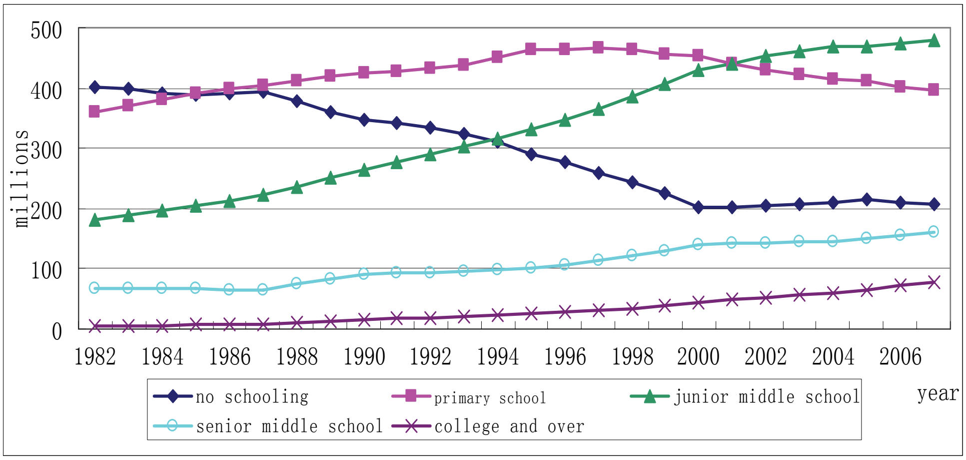


Figure III.1.3 shows population by educational attainment from 1982 to 2007. The illiterate population was cut in half from 402 million in 1982 to 201 million in 2000, but was relatively stable from 2000 to 2007. The number of primary school graduates increased from 359 million in 1982 to the peak of 466 million in 1997, then declined gradually to 399 million in 2007. This decline is expected as more primary school graduates continue on to higher education level instead of terminating formal education. This is also evident in the rapid growth of junior middle school graduates.

Junior middle school students registered the largest growth among all education levels: the number of junior middle school graduates increased from 181 million in 1982 to 471 million in 2007. This might be related to the implementation of 9-Year Compulsory Schooling since 1994 (9-year schooling amounts to completing junior middle school). However, the growth slowed after 2001. Senior middle school and college and over, both started from very low numbers and have grown significantly. Senior middle school graduates increased from 68 million in 1982 to 166 million in 2007, while college and above increased from only 6 million in 1982 to 76 million in 2007.

Figure III.1.4 Population of different educational levels by gender, 1982

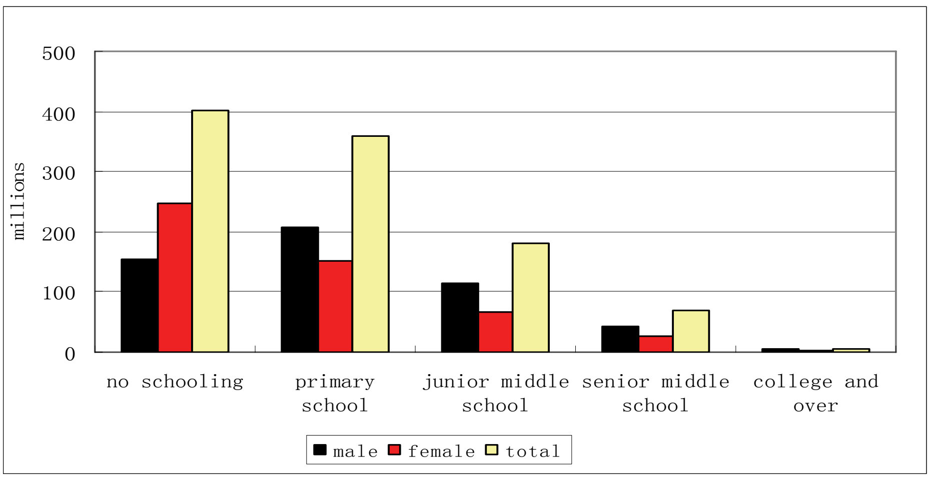


Figure III.1.5 Population of different educational levels by gender, 1988

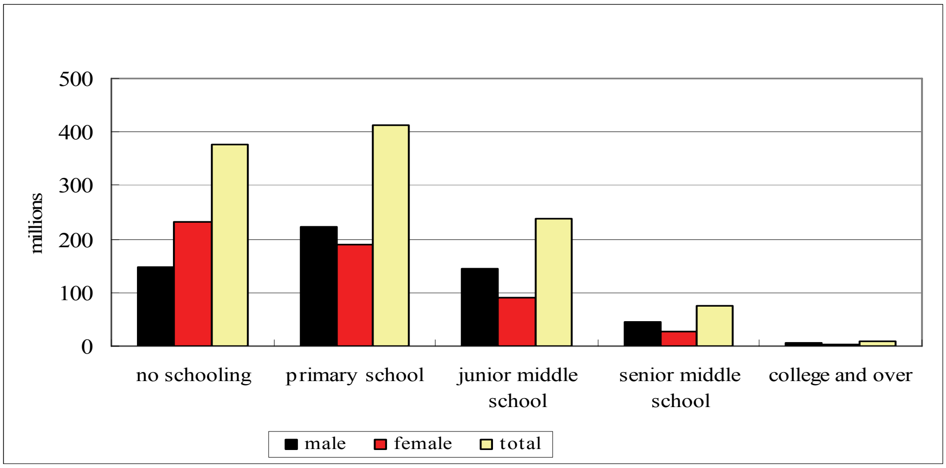


Figure III.1.6 Population of different educational levels by gender, 1998

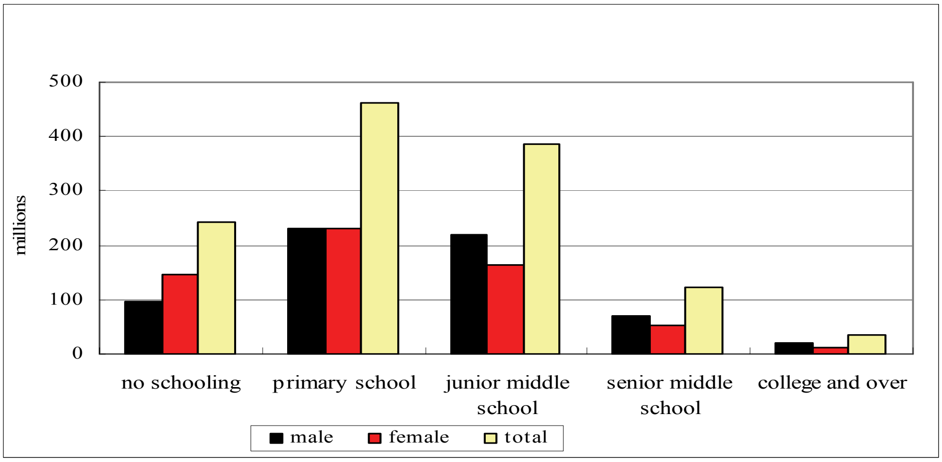
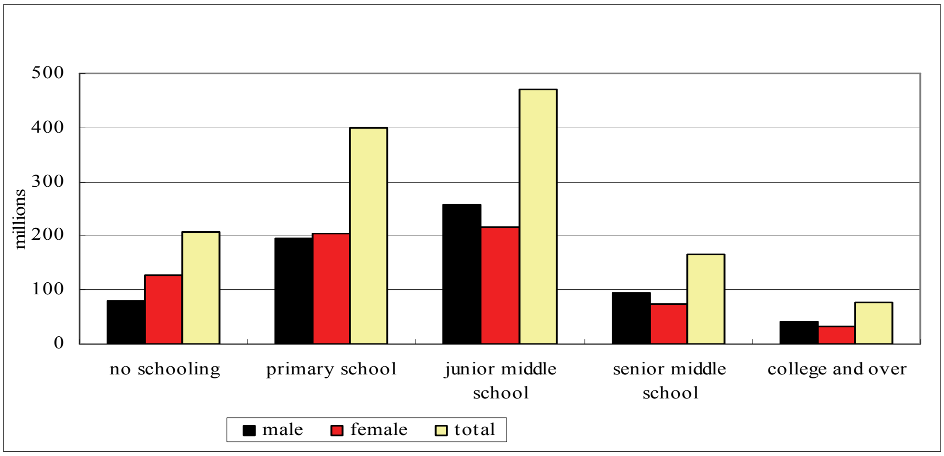


Figure III.1.7 Population of different educational levels by gender, 2007



We next take a closer look at the changes in the distribution of education attainment in the population from 1982 to 2007. Figures III.1.4~7 show the rightward shift of the educational attainment distribution in the population. In 1982, among the five education levels, the illiterates take up the largest portion. The 1988 distribution is dominated by people with primary and less education, i.e. the distribution remains heavily skewed to the right. In 1998, the distribution is dominated by primary and junior middle graduates. By 2007, junior middle has become the dominant education level. The distribution is still skewed to the right, but it is much less so than in 1982. Moreover, female educational attainment has improved more relative to that of males; the number of illiterate females decreased faster than that of illiterate males, while the gender differences at higher education levels shrunk considerably. As a result, the female educational attainment distribution is becoming similar to that of the male, despite the drastic difference in 1982.

Figure III.1.8 Population of different educational levels by urban and rural, 1982

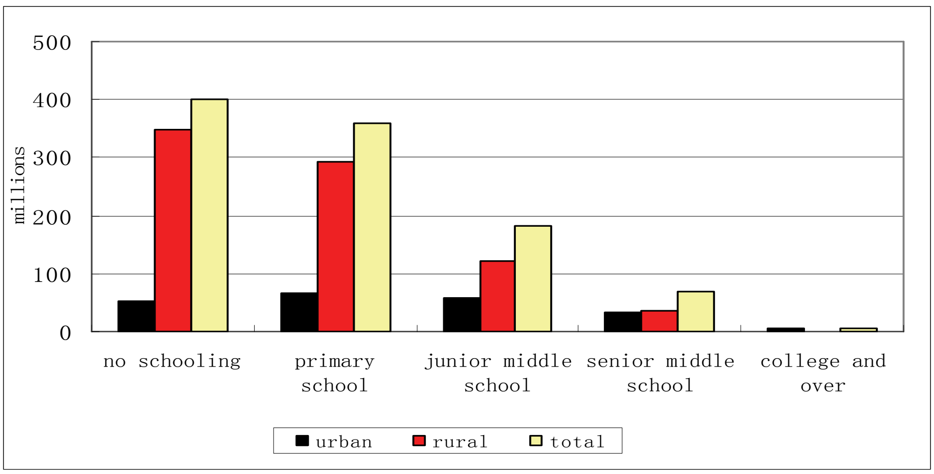


Figure III.1.9 Population of different educational levels by urban and rural, 1988

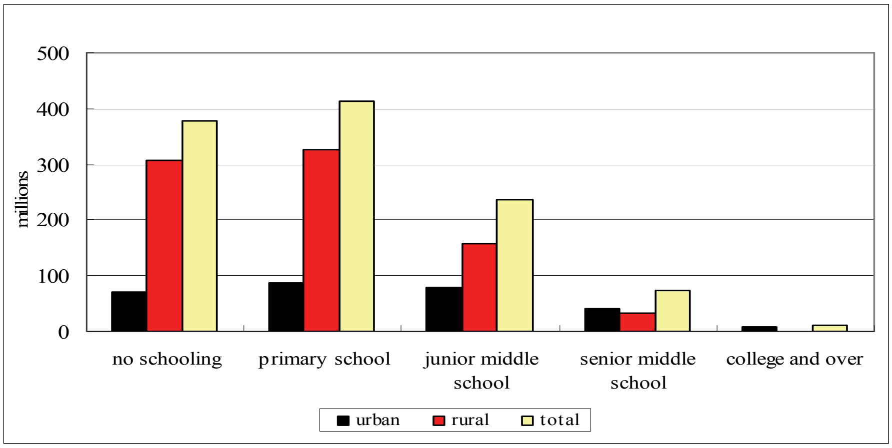


Figure III.1.10 Population of different educational levels by urban and rural, 1998

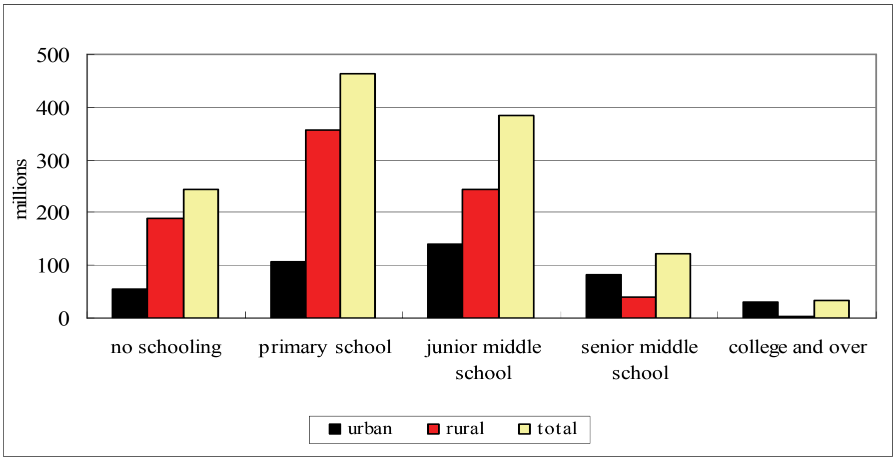
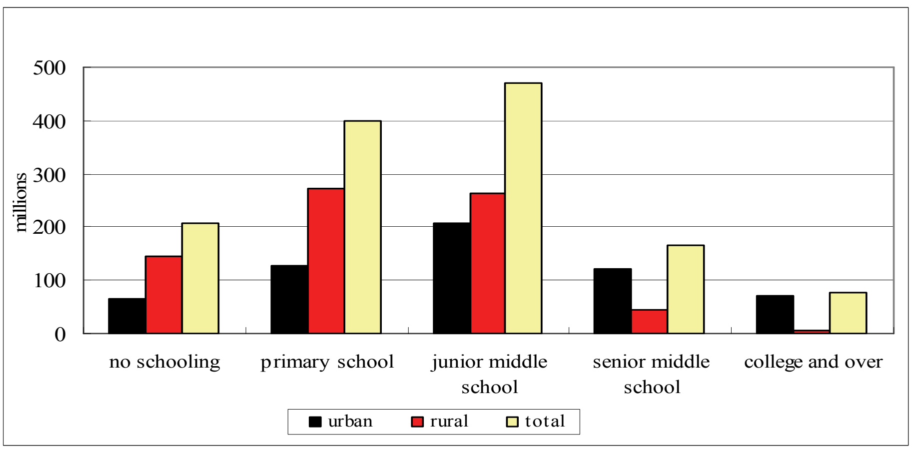


Figure III.1.11 Population of different educational levels by urban and rural, 2007



Figures III.1.8~11 disaggregate the data into rural and urban samples. Not surprisingly, most of the illiterate population resided in the rural area. However, the rural illiterate population fell from 349 million in 1982 to 144 million in 2007. Although the urban illiterate population changed slightly in absolute terms, its share in the urban population fell from nearly a quarter in 1982 to 10.86% in 2007. In the meantime, in the highest three levels of education (junior middle, senior middle, and college and over), urban growth outpaced rural growth. For example, the urban junior middle school population more than tripled from 58 million to 208 million, while the rural junior middle school population roughly doubled, from 123 million to 263 million. The comparison is more startling in the highest two education levels. The urban senior middle school population increased from 18 million to 122 million, while the rural senior middle school population only increased from 35 million to 44 million. The urban college and over population increased 14-fold, from 5 million to 71 million, while in rural areas, it grew 6-fold, but remained very small, at only 5 million individuals.

Note that during the entire sample period, the rural population far exceeded the urban population. Although both the urban and the rural distributions have improved, i.e. less skewed to the right, the improvement has certainly been more rapid and obvious in the urban area. One caveat, however, is that the result might be caused by better educated people migrating from rural to urban areas. We take special measures to control for that effect (See Appendix A).

III.2 Obtaining parameter estimates of the Mincer equation

One important component of the income approach is the estimation of future potential earnings for all individuals in the population. We conduct estimation and make projection based on the basic Mincer (1974) equation. It has been shown that there are significant differences in the structure of the earnings equation across gender and between the rural and urban population. To ensure our income estimates to be as accurate as possible, we estimate the parameters for the rural and urban population by gender and year using survey data in selected years and derive their imputed values for missing years over the period of 1985 to 2020.

We first estimate the basic Mincer equation:

$$\ln(inc) = \alpha + \beta \cdot e + \gamma \cdot exp + \delta \cdot exp^2 + u \qquad (1)$$

where ln(*inc*) is the logarithm of earnings, *e* is years of schooling, e*xp* and exp^2 are, respectively, years of work experience and experience squared, and *u* is a random error. The coefficient *α* is an estimate of the average log earnings of individuals with zero years of schooling and work experience, *β* is an estimate of the return to an extra year of schooling, and *γ* and *δ* measure the return to investment in on-the-job training.

Equation (1) has been the workhorse widely adopted in empirical research on earnings determination. It has been estimated on a large number of data sets for numerous countries and time periods. Many studies have applied the model to Chinese data and found evidence consistent with the human capital theory. Notable studies include, among others, Liu (1998), Maurer-Fazio (1999), Li (2003), Fleisher and Wang (2004), Yang (2005), and Zhang *et al.* (2005). Following the convention of a large body of empirical literature, we estimate equation (1) by ordinary least squares.[9]

The data used for estimating the parameters of the earnings equation come from two well-known household surveys in China. The first is the annual Urban Household Survey (UHS) conducted by the National

[9] Griliches (1977) finds that accounting for the endogeneity of schooling and ability bias does not alter the estimates of earnings equation. Ashenfelter and Krueger (1994) also conclude that omitted ability variables do not cause an upward bias in the estimated parameters of equation (1).

Statistical Bureau of China over the period of 1986-1997. We use this data set to estimate the parameters of equation (1) for each gender of the urban population by year, and then extract fitted estimates by applying linear or exponential time trends. We use the fitted time trends to generate the imputed parameters of the earnings equation for the urban population for the period 1985 through 2020.

The second data set we use is the China Health and Nutrition Survey (CHNS) for the years of 1989, 1991, 1993, 1997, and 2000. This survey covers both the urban and rural population. We use CHNS to obtain earnings-equation parameter estimates by year for each gender and separately for the rural and urban population. We calculate the urban-to-rural ratio for each of these parameters. We then use the ratio to fit a time trend model (i.e. interpolate and extrapolate), which is used to generate fitted values of the urban-to-rural ratio over the period 1985 to 2020. We use the fitted ratios along with the imputed parameters for the urban population to derive the imputed parameters for the rural population over the period 1985 to 2020.

III.2.1 Imputing the earnings equation parameters for the urban population

The UHS is a representative sample of the urban population. The sample size varies from year to year, ranging from a low of 4,934 respondents in 1986 to a high of 31,266 respondents in 1992. Individual earnings are annual wage incomes, which include basic wage, bonus, subsidies and other work-related incomes. Years of schooling are calculated using the information on the level of schooling completed: primary school equals 6 years of schooling, junior middle school 9 years, senior middle school 12 years, professional school 11 years, community college 15 years, and college and above 16 years. Assuming schooling begins at age 6, we approximate work experience by age minus years of schooling minus 6. As the minimum legal working age is 16 and the retirement ages are 60 and 55 for males and females respectively, we restrict our sample to include individuals who are currently employed and are between 16 and 60 years of age for male workers and between 16 and 55 for female workers. Self-employed and temporary job holders are excluded, so are those who failed to report wage income or educational attainment. Table B.2 in Appendix B contains means and standard deviations of these variables.

We use the UHS data to estimate the earnings equation for each gender by year. The estimates are reported in Table B.3 in Appendix B. They are by and large in line with the estimates reported in previous studies using the same or similar Chinese data. The constant term, which measures the base wage for the no-school no-experience population, clearly reveals the male advantage (Figure III.2.1.1). Returns to schooling are positive and in general increasing over the sample years (Figure III.2.1.2). Male return increased from a meager 1.7% in 1986 to 7.2% in 1997, while female return also increased from 4.2% in 1986 to 10.8% in 1997. Wang, Fleisher, Li, and Li (2009) also reports that female rates of return dominate male returns, and they offered an explanation. Rising returns to education have been a ubiquitous phenomenon in transitional economies when the Soviet-type wage grid was replaced by market wages (Fleisher, Sabirianova, Wang 2005). Earnings also increase with work experience but at a decreasing rate — a pattern found in most studies. Over time the earnings-experience profile shifts up for male (Figure III.2.1.3) but fluctuates for females. For most recent years the male profile doesn't curve downward as much as that of the female (Figure III.2.1.4), and the male profile is much higher than the female profile, indicating uniformly higher return to experience for male than for female, *ceteris paribus*.

Figure III.2.1.1 Constant term, zero-education zero-experience, UHS samples

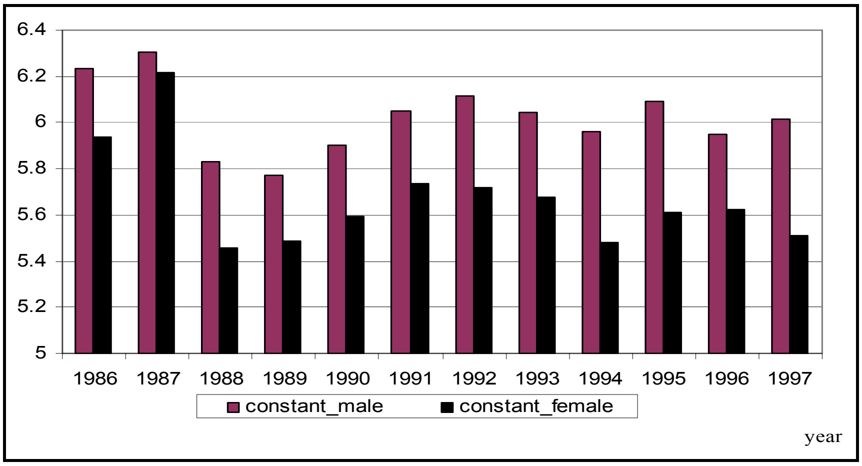


Figure III.2.1.2 Return to education, UHS samples

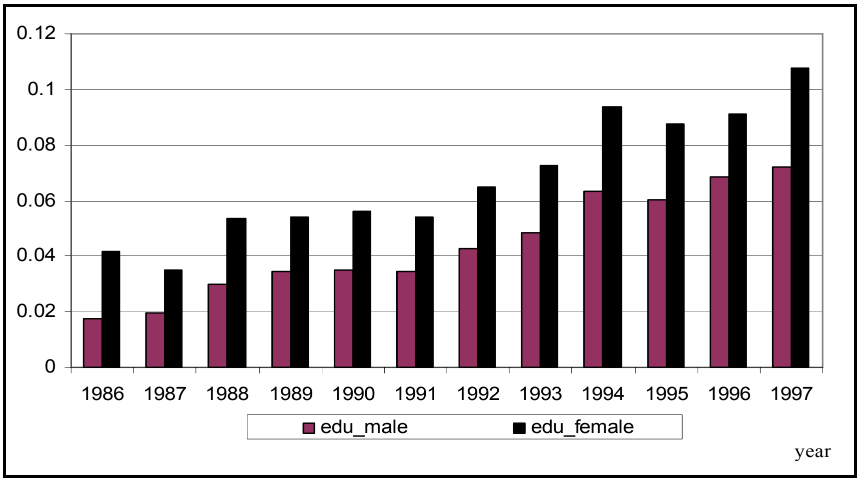


Figure III.2.1.3 Return to experience, male, UHS samples

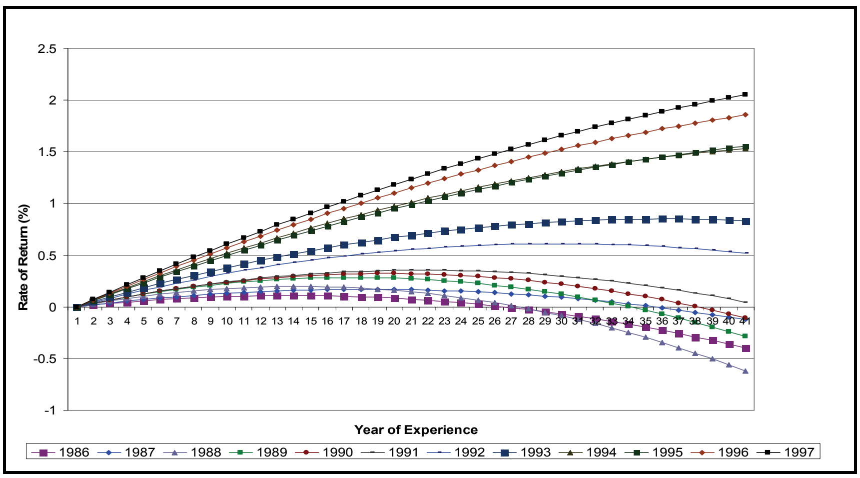
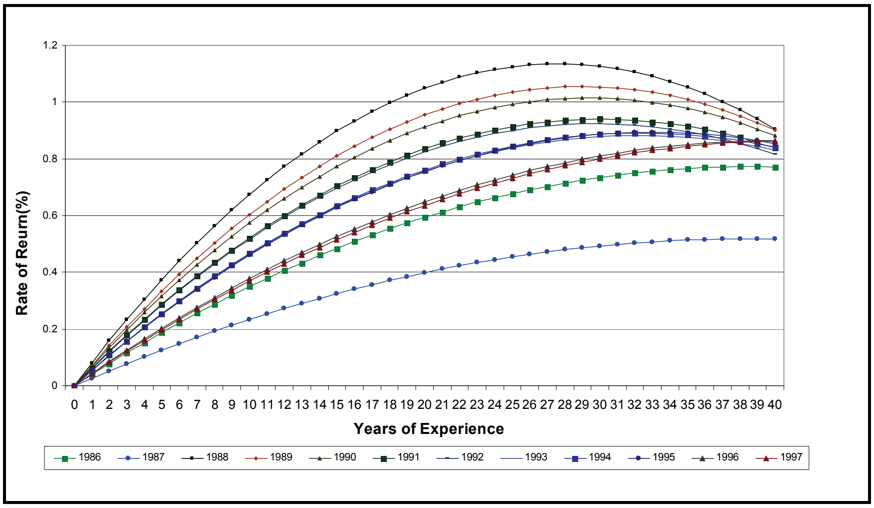


Figure III.2.1.4 Return to experience, female, UHS samples



Figures B.1~8 in Appendix B show that when we plot each of the parameter estimates against time, they are generally trended. The large changes in the values of the estimated intercepts and coefficients on work experience and experience squared from 1986 and 1987 to 1988 are puzzling. We suspect that these changes may have been due to inconsistency in survey methodology adopted across the initial few years of the survey. We exclude these outliers from the time trend estimation of the parameters. For each parameter, we regress its estimates reported in Table B.3 (Appendix B) against time under two alternate specifications: a linear time trend model, and an exponential trend model where the logarithm of the parameter estimate is the dependent variable. The AIC values, a popular test for model selection, suggest that the linear time trend specification is appropriate for the intercept and the schooling parameter, while the exponential trend specification is suitable for the parameters associated with work experience and experience squared. As the coefficient on experience squared is negative, the dependent variable is defined as the log of the absolute value of the parameter estimates. Figures B.9~16 in Appendix B show the actual parameter estimates over the period 1988 to 1997 along with the fitted trend lines.

We use the fitted trend lines to generate imputed values of the parameters for each gender by year over the period 1985 to 2007. While there is some evidence that the pre-1997 trends of these parameters, particularly the one associated with schooling, continued after 1997 and up to 2007 (see e.g., Zhang *et. al.* 2005), it is unclear if the trends will extend beyond 2007. We therefore assume, probably rather conservatively, that the earnings equation parameters remain constant for the period 2007 to 2020 and are equal to the fitted values of their counterparts in 2007. Table III.2.1.1 reports the imputed values of the parameters for the urban population by gender and year.

Table III.2.1.1: Imputed earnings equation parameters for the urban population,1985 to 2020

year	male				female			
	α	β	γ	δ	α	β	γ	δ
1985	5.81248	0.01089	0.08555	-0.00147	5.55553	0.02677	0.09859	-0.00209
1986	5.83390	0.01595	0.08061	-0.00134	5.56000	0.03301	0.09198	-0.00187
1987	5.85532	0.02101	0.07595	-0.00122	5.56447	0.03926	0.08581	-0.00167
1988	5.87673	0.02608	0.07156	-0.00111	5.56894	0.04550	0.08006	-0.00150
1989	5.89815	0.03114	0.06742	-0.00102	5.57342	0.05174	0.07469	-0.00134
1990	5.91956	0.03620	0.06353	-0.00093	5.57789	0.05798	0.06968	-0.00120
1991	5.94098	0.04126	0.05986	-0.00084	5.58236	0.06422	0.06501	-0.00107
1992	5.96239	0.04632	0.05640	-0.00077	5.58683	0.07046	0.06065	-0.00096
1993	5.98381	0.05138	0.05314	-0.00070	5.59130	0.07670	0.05658	-0.00086
1994	6.00522	0.05645	0.05007	-0.00064	5.59577	0.08295	0.05279	-0.00077
1995	6.02664	0.06151	0.04717	-0.00058	5.60024	0.08919	0.04925	-0.00069
1996	6.04805	0.06657	0.04445	-0.00053	5.60472	0.09543	0.04595	-0.00062
1997	6.06947	0.07163	0.04188	-0.00048	5.60919	0.10167	0.04287	-0.00055
1998	6.09088	0.07669	0.03946	-0.00044	5.61366	0.10791	0.03999	-0.00049
1999	6.11230	0.08176	0.03718	-0.00040	5.61813	0.11415	0.03731	-0.00044
2000	6.13372	0.08682	0.03503	-0.00037	5.62260	0.12040	0.03481	-0.00040
2001	6.15513	0.09188	0.03300	-0.00033	5.62707	0.12664	0.03248	-0.00035
2002	6.17655	0.09694	0.03110	-0.00030	5.63155	0.13288	0.03030	-0.00032
2003	6.19796	0.10200	0.02930	-0.00028	5.63602	0.13912	0.02827	-0.00028
2004	6.21938	0.10707	0.02761	-0.00025	5.64049	0.14536	0.02637	-0.00025
2005	6.24079	0.11213	0.02601	-0.00023	5.64496	0.15160	0.02460	-0.00023
2006	6.26221	0.11719	0.02451	-0.00021	5.64943	0.15785	0.02295	-0.00020
2007	6.28362	0.12225	0.02309	-0.00019	5.65390	0.16409	0.02141	-0.00018
2008	6.28362	0.12225	0.02309	-0.00019	5.65390	0.16409	0.02141	-0.00018
2009	6.28362	0.12225	0.02309	-0.00019	5.65390	0.16409	0.02141	-0.00018
2010	6.28362	0.12225	0.02309	-0.00019	5.65390	0.16409	0.02141	-0.00018
2011	6.28362	0.12225	0.02309	-0.00019	5.65390	0.16409	0.02141	-0.00018
2012	6.28362	0.12225	0.02309	-0.00019	5.65390	0.16409	0.02141	-0.00018
2013	6.28362	0.12225	0.02309	-0.00019	5.65390	0.16409	0.02141	-0.00018
2014	6.28362	0.12225	0.02309	-0.00019	5.65390	0.16409	0.02141	-0.00018
2015	6.28362	0.12225	0.02309	-0.00019	5.65390	0.16409	0.02141	-0.00018
2016	6.28362	0.12225	0.02309	-0.00019	5.65390	0.16409	0.02141	-0.00018
2017	6.28362	0.12225	0.02309	-0.00019	5.65390	0.16409	0.02141	-0.00018
2018	6.28362	0.12225	0.02309	-0.00019	5.65390	0.16409	0.02141	-0.00018
2019	6.28362	0.12225	0.02309	-0.00019	5.65390	0.16409	0.02141	-0.00018
2020	6.28362	0.12225	0.02309	-0.00019	5.65390	0.16409	0.02141	-0.00018

III.2.2 Imputing the earnings equation parameters for the rural population

The CHNS is an ongoing international collaborative project between the Carolina Population Center at the University of North Carolina at Chapel Hill and the National Institute of Nutrition and Food Safety at the Chinese Center for Disease Control and Prevention and was designed for evaluating the impact of social and economic transformation of the Chinese society on socioeconomic, demographic, and health behaviors of the urban and rural population. The survey also contains information on income, age and educational attainment, which we use to estimate the earnings equation by year for each gender and separately for the urban and rural population. For the urban sample, earnings contain wage income and subsidies from work.

The rural sample contains only household income, which includes family members' incomes from the collective or household productions or both in five distinct activities: gardening, farming, raising livestock, fishing, and small handicraft and family businesses. We allocate household income to each individual member according to his or her working hours as a share of the household's total. Years of schooling are calculated based on the reported grade or years completed (depending on the sample year). Work experience is approximated by age minus years of schooling minus 6. We restrict our sample to males between 16 and 60 years of age and females between 16 and 55 who reported information on education and income. Table B.5 in Appendix B contains the summary statistics of the variables used.

We use the CHNS data to estimate equation (1) by gender and separately for the rural and urban samples for each of the sample year (i.e., 1989, 1991, 1993, 1997, and 2000). The parameter estimates, which are reported in Table B.6 in Appendix B, are then used to calculate the urban-to-rural ratio for each parameter by gender. We use the ratios to fit an exponential trend model, which is used to generate the fitted ratios for the period 1985 to 2007. We assume that the ratios remain constant for the period 2007 to 2020 and are equal to the fitted values of their counterparts in 1997. Table B.7 in Appendix B reports the fitted ratios. These fitted urban-to-rural ratios by themselves provide interesting insights. For example, in 1985, the urban no-schooling no-experience male cohort was on average paid 9.8% more than its rural counterpart, and by 2007 this gap has increased to 14.6%. In the meantime, the urban no-schooling no-experience female cohort was on average paid 6.7% more than its rural counterpart, and by 2007 the rural cohort was paid 1.8% more than the urban cohort. Return to education is always higher for rural male than for urban male. In 1985, the rate of return was 16% higher for rural male, and by 2007 it was 33% higher. For female, however, it is a different story. Return to education for urban female was 63% higher than rural female, but by 2007 the return to urban female was 22% less than rural female. The relation between urban and rural return to experience has also changed. All of these are not central to our current project, but nevertheless deserves attention in future research.

We use these ratios along with the imputed parameters for the urban population in Table III.2.1.1 to impute parameters for the rural population which are presented in Table III.2.2.1.

Table III.2.2.1: Imputed earnings equation parameters for the rural population, 1985 to 2020

year	male				female			
	α	***β***	***γ***	***δ***	***α***	***β***	***γ***	***δ***
1985	5.29358	0.01297	0.06773	-0.00093	5.20888	0.01646	0.12262	-0.00258
1986	5.30279	0.01919	0.06613	-0.00090	5.23264	0.02099	0.10967	-0.00219
1987	5.31194	0.02554	0.06456	-0.00088	5.25651	0.02580	0.09809	-0.00186
1988	5.32103	0.03201	0.06303	-0.00085	5.28047	0.03092	0.08773	-0.00157
1989	5.33007	0.03860	0.06154	-0.00083	5.30455	0.03635	0.07846	-0.00133
1990	5.33906	0.04532	0.06008	-0.00080	5.32873	0.04212	0.07017	-0.00113
1991	5.34799	0.05218	0.05866	-0.00078	5.35302	0.04823	0.06276	-0.00096
1992	5.35687	0.05916	0.05727	-0.00076	5.37741	0.05472	0.05613	-0.00081
1993	5.36569	0.06628	0.05591	-0.00074	5.40191	0.06158	0.05020	-0.00069
1994	5.37446	0.07354	0.05459	-0.00071	5.42653	0.06885	0.04490	-0.00058
1995	5.38317	0.08094	0.05330	-0.00069	5.45125	0.07654	0.04016	-0.00049
1996	5.39183	0.08847	0.05204	-0.00067	5.47607	0.08468	0.03592	-0.00042
1997	5.40043	0.09615	0.05080	-0.00066	5.50101	0.09327	0.03212	-0.00035
1998	5.40899	0.10397	0.04960	-0.00064	5.52606	0.10236	0.02873	-0.00030
1999	5.41748	0.11194	0.04843	-0.00062	5.55122	0.11195	0.02569	-0.00025
2000	5.42593	0.12005	0.04728	-0.00060	5.57649	0.12207	0.02298	-0.00022
2001	5.43432	0.12832	0.04616	-0.00058	5.60187	0.13276	0.02055	-0.00018
2002	5.44266	0.13674	0.04507	-0.00057	5.62736	0.14402	0.01838	-0.00015
2003	5.45095	0.14532	0.04400	-0.00055	5.65297	0.15590	0.01644	-0.00013
2004	5.45918	0.15405	0.04296	-0.00054	5.67869	0.16842	0.01470	-0.00011
2005	5.46736	0.16295	0.04194	-0.00052	5.70452	0.18161	0.01315	-0.00009
2006	5.47549	0.17200	0.04095	-0.00051	5.73047	0.19549	0.01176	-0.00008
2007	5.48357	0.18122	0.03998	-0.00049	5.75653	0.21012	0.01052	-0.00007
2008	5.48357	0.18122	0.03998	-0.00049	5.75653	0.21012	0.01052	-0.00007
2009	5.48357	0.18122	0.03998	-0.00049	5.75653	0.21012	0.01052	-0.00007
2010	5.48357	0.18122	0.03998	-0.00049	5.75653	0.21012	0.01052	-0.00007
2011	5.48357	0.18122	0.03998	-0.00049	5.75653	0.21012	0.01052	-0.00007
2012	5.48357	0.18122	0.03998	-0.00049	5.75653	0.21012	0.01052	-0.00007
2013	5.48357	0.18122	0.03998	-0.00049	5.75653	0.21012	0.01052	-0.00007
2014	5.48357	0.18122	0.03998	-0.00049	5.75653	0.21012	0.01052	-0.00007
2015	5.48357	0.18122	0.03998	-0.00049	5.75653	0.21012	0.01052	-0.00007
2016	5.48357	0.18122	0.03998	-0.00049	5.75653	0.21012	0.01052	-0.00007
2017	5.48357	0.18122	0.03998	-0.00049	5.75653	0.21012	0.01052	-0.00007
2018	5.48357	0.18122	0.03998	-0.00049	5.75653	0.21012	0.01052	-0.00007
2019	5.48357	0.18122	0.03998	-0.00049	5.75653	0.21012	0.01052	-0.00007
2020	5.48357	0.18122	0.03998	-0.00049	5.75653	0.21012	0.01052	-0.00007

III.3 Growth rates of real income and the discount rate

To measure lifetime earnings for all individuals in the population, we need to project incomes for future years, discount these incomes back to the present, and weight income for each individual by the age- and gender-specific probability of survival. We use the imputed earnings equation parameters to estimate earnings for all individuals in a given year, and then derive earnings for future years until retirement assuming real earnings grow at a constant rate.[10] The main task of this section is to estimate the expected growth rate of real income and select an appropriate discount rate. Since the real income grew at fairly different rates in the past for the urban and rural population, we estimate them separately.

III.3.1 Growth rates of real income

Assuming that the technology is labor-augmenting, we specify the aggregate production function as:

$$Y = (AL)^a K^b$$

where Y is output, A denotes a technology factor, L denotes labor input, and K physical capital input. The average product of labor or labor productivity is proportional to the marginal product of labor.[11] Because the marginal product of labor equals the real wage when the labor market is in equilibrium, labor productivity and the real wage are expected to grow at the same rate. Therefore, the growth rate of real output per employed worker can serve as a reasonable estimate for the growth rate of the real wage.

---

[10] Mincer equation parameter estimates are used to calculate the cohort-wise labor income for a given year, it is not used to project future income.
[11] The marginal product of labor is given by $\beta Q/L$, where Q/L is the average product of labor.

National Statistical Bureau of China publishes nominal GDP and real GDP index (in 1978 prices) by sector (primary industry, secondary industry, and tertiary industry). We derive real GDP as the product of nominal GDP in the base year and real GDP index. The labor productivity in the rural sector is defined as real GDP of the primary industry divided by the number of persons employed in the primary industry. The labor productivity in the urban sector is the ratio of real GDP of the secondary and tertiary industries to the number of persons employed in these industries.

Table D.2 in Appendix D presents the annual growth rate of labor productivity along with the GDP and employment data over the period 1978 to 2007. These numbers indicate that in the past 30 years labor productivity grew on average 4.11% and 6% per annum in the rural and urban sectors, respectively. We assume labor productivities (and hence the real income) continue to grow annually at these average rates.[12]

III.3.2 The discount rate

The discount rate that is used to value future incomes in present terms should reflect the rate of return one expects from investments over a long time horizon. In this regard, the interest rate paid on government bonds is a good proxy. We choose a discount rate of 3.14%, which is the average interest rate on the 10-year government bonds issued to individual investors over the period 1996 to 2007, net of the average rate of inflation over the same period. It should be noted that our discount rate is lower than the discount rates used in the Jorgenson and Fraumeni studies cited in this report.

[12] One obvious concern is how fast these rates will converge to the long-run steady-state rates, and what are the long-run steady-state rates. Our future research will address these issues.

III.4 Additional data imputations and assumptions for the Jorgenson- Fraumeni estimates

Besides annual population data by age, sex, and educational attainment, the Jorgenson-Fraumeni method requires additional information on the lifetime income, enrollment rate, growth rate of real wage, and discount rate. We briefly discuss how we construct these supplemental data sets in this section. Some parameters have to be set at values appropriate for China. Detailed information can be found in the appendixes.

Following Jorgenson and Fraumeni, an individual may assume one of the following six statuses at any time: no school or work (age 0-5), school only (age 6-16), work and school (age 16 to age), work only (age to retirement), and retirement (age 60+ for male and 55+ for female). Each status implies a different pattern of age-income profile, therefore the method of computing lifetime income shall be different.

We first estimate a standard Mincer equation (i.e. with a regression of annual income on schooling years, work experience, and work experience squared) with microeconomic data sets (China Household Income Project, China Health and Nutrition Survey, and Urban Household Survey). We use annual employment rates by age, sex, and educational attainment (from China Population Statistical Yearbook and China Population Census) to convert annual income into annual market income. Then the lifetime income for each age/sex/education category can be calculated using the methodology described in the earlier section.

For the in-school population, we carefully derive the number of people in each education level with data on new enrollment, mortality rate, and attrition rate. We consider the following five categories of schooling: no schooling, primary school, junior middle school, senior middle school, and college and above or for six categories of schooling college and university and above. We compute lifetime income for every grade at each education level, taking into account how likely the individual will continue into the next grade and the next education level. For the five categories of schooling estimates college and above is the highest education level. For the six categories of schooling estimates college or university and above are the highest education levels. We do not allow for the possibility that one can go to college then followed by university.

As not all data is available by single year of age or by individual level of education, some additional imputations and assumptions are needed. Imputations having to do with creating data sets by single year of age and initial age of enrollment are described in Appendix A. Enrollment and grade advancement imputations and assumptions are described in this section.

The imputation of two components of the J-F human capital estimates is described in this section: 1) Number of years until an education category is completed, and 2) The probability of advancing to the next higher education category. A decision was made to assume that all students complete a grade level (if they continue) in the same number of years: 6 for primary, 3 for junior middle, and 3 for senior middle school. It is also assumed that no drop-outs return to school and that education continues without a break. These assumptions are also made by J-F. The probability of advancing to the next higher education level is estimated as the average ratio of the sum of all students of any age in a year who are initially enrolled to the sum of all students of any age initially enrolled in the next higher education level "X" years later. "X" depends upon the number of years it takes to complete an education level. The imputations and assumptions allow for the appropriate discounting of a future higher income level.

In each case, advancing students are tracked from their age of initial enrollment, through individual grade levels, until they advance to the next higher level. The number of years discounted until they realize the higher level of lifetime income depends on the number of years it takes to advance given the current grade of enrollment.

Then, we treat the terminal education level as a probabilistic event, and therefore the lifetime income is a forecast based on the contemporary information set, except that the probability of advancing depends on initial enrollments at a higher education level in subsequent years. For instance, the lifetime income of a student who is in the first year of junior middle school, assuming she will live to finish junior middle school and goes onto senior middle school depends upon an adjusted lifetime income of someone who is currently three years older and whose educational attainment is senior middle school. The adjustments include those for three years of labor income (wage) growth and three years of discounting,

$$mi_{s,a,Grade1of\quad junior} = mi_{s,a+3,Grade1of\quad senior} \times senr_{s,a,enr} \times senr_{s,a+1,enr} \times senr_{s,a+2,enr}$$
$$\times sr_{s,a+1} \times sr_{s,a+2} \times sr_{s,a+3} \times \left(\frac{real\quad income\quad growth\quad rate}{discount\quad rate} \right)^3$$

We use the average labor productivity growth rate as the real income (wage) growth rate. Moreover, we use the labor productivity growth rate in the primary sector as the rural real wage growth rate, and labor productivity growth rate in the secondary and tertiary sectors as the urban real wage growth rate. For our sample period of 1985-2007, it is 6% for urban workers and 4.11% for rural workers. As of the subjective discount rate as noted earlier, we use the long-term government bonds (average real) interest rate for the sample period, and it is 3.14%.

IV Result discussions

IV.1 Total human capital stock, GDP, and physical capital stock

Our main results are based on the J-F approach. The estimated total human capital stock at the national level for 1985-2007 is reported in Table IV.1.1. Columns 1 and 2 contain the total human capital measured in nominal terms, and columns 3 and 4 present the total human capital measured in real terms (in 1985 RMB). In this table, the real values are calculated using CPI.[13] Figure IV.1.1 shows the trend of human capital in both real and nominal values.

Before 2000, five education categories were reported by the National Bureau of Statistics of China. They are: no school, elementary school, junior middle school, senior middle school, and college and above. Starting from 2000, the college and above was further divided into two categories: three-year college, and four-year college and above.[14] To take advantage of this more detailed information on educational attainment, we create a separate human capital series starting from 2000. As can be seen from Figure IV.1.2, total human capital becomes larger with six education categories. This is because the lifetime incomes of graduates of four-year college and above are higher than those who graduated from three-year colleges.

[13] Because the total human capital is the sum of rural and urban human capital, we use CPI for rural and urban separately in the estimation.

[14] When we estimate Mincer equation to generate annual earnings, we assign 15 years of schooling for the category of three-year college; and assign 16 years of schooling for the category four-year college and above. Because we use the lower bound of schooling for this education category, the amount of human capital is underestimated.

Table IV.1.1 Nominal and real human capital, nominal GDP (1985 as base year for real series, in trillions)

year	nominal human capital		real human capital		nominal GDP	ratio of human capital to GDP (current prices)
	five education categories	six education categories	five education categories	six education categories		
1985	26.98		26.98		0.90	29.92
1986	29.85		28.03		1.03	29.05
1987	33.59		29.38		1.21	27.85
1988	41.64		30.61		1.50	27.68
1989	50.82		31.68		1.70	29.91
1990	54.57		33.02		1.87	29.23
1991	59.35		34.65		2.18	27.25
1992	66.63		36.47		2.69	24.75
1993	82.96		39.48		3.53	23.48
1994	111.63		42.73		4.82	23.16
1995	136.58		44.61		6.08	22.47
1996	165.55		49.76		7.12	23.26
1997	192.18		56.01		7.90	24.33
1998	206.34		60.48		8.44	24.45
1999	224.15		66.46		8.97	25.00
2000	245.00	249.64	72.19	73.50	9.92	24.69
2001	263.75	269.02	77.05	78.52	10.97	24.05
2002	281.04	287.23	82.63	84.38	12.03	23.36
2003	307.23	314.71	89.20	91.29	13.58	22.62
2004	338.20	346.73	94.59	96.90	15.99	21.15
2005	370.45	380.48	101.78	104.46	18.32	20.22
2006	404.46	416.40	109.46	112.60	21.19	19.08
2007	459.82	474.23	118.75	122.38	24.95	18.43

Figure IV.1.1 Nominal and real human capital, 1985-2007

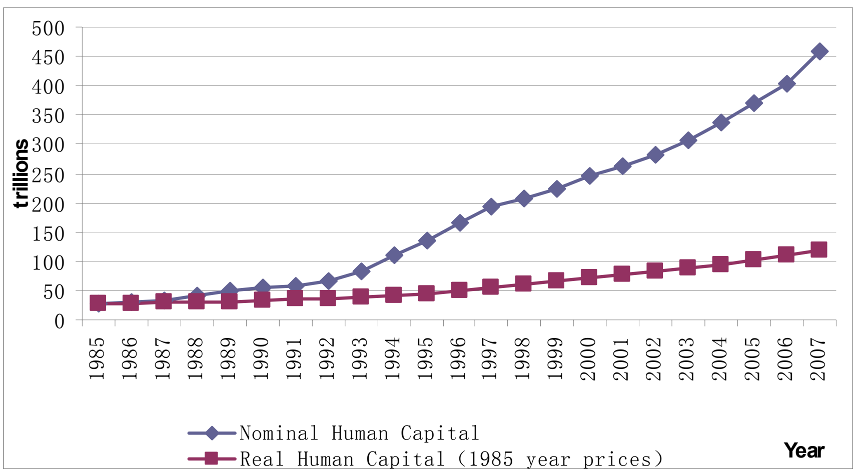


Figure IV.1.2 Real total human capital by different education categories, 2000-2007

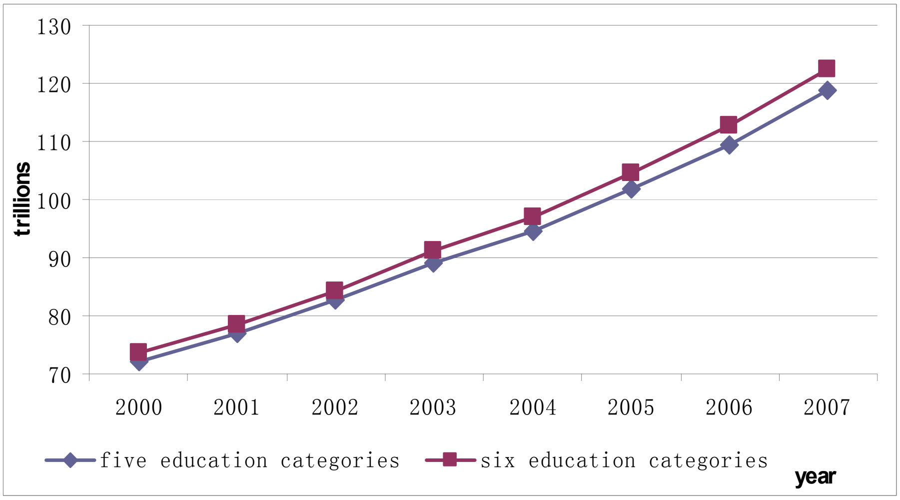


Table IV.1.2 International comparison of human capital estimates

	Canada 2007	Norway 2006	New Zealand 2001	U.S. 2006	Australia 2001	China 2001	China 2006	China 2007
	2007 USD	2006 USD	2001 USD	2006 USD	2001 USD	2001 USD	2006 USD	2007 USD
Age Range	15-74	15-67	25-65	0-80	18-65	male 0-60, female 0-55		
Per capita human capital	607,696	-	145,967	over 700,000	-	28,383	45,454	54,213
Total human capital (trillions)	15.08	2.38	0.29	212	3.62	31.87	50.73	60.47
Ratio of human capital to GDP	11	8	6	over 15	10	24	19	18

In order to get a sense of the magnitude of the estimated total human capital in China, we also reported nominal GDP in Table IV.1.1. The ratio of estimated (market) human capital to GDP generally declines over time until 2005-7, when it is between 18 and 20. Jrogenson and Fraumeni (1992a)'s estimates of the ratio of total market human capital to GDP in the U.S. from 1947 to 1986 is between 18 and 22. A summary of international comparison of human capital estimates is reported in Table IV.1.2. China's total human capital is quite large, more than any country except the U.S. However, China's per capita human capital is still very small. In China during the later period, the growth of population slowed but the economy continues to grow at a higher rate, which contributes to the declining ratio of human capital to GDP (Figure IV.1.3).

Figure IV.1.3 Ratio of nominal total human capital and nominal GDP

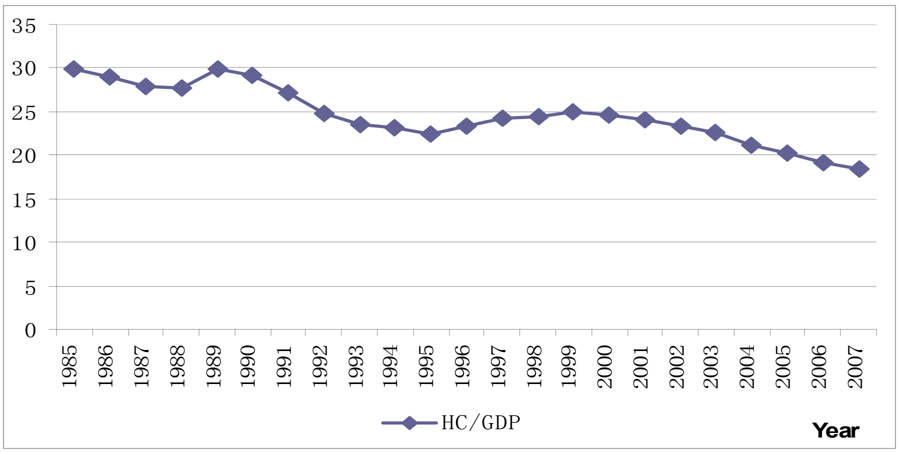


Table IV.1.3 Total human capital and physical capital (Zhang *et. al.* 2004), 1985-2000, in trillions

year	total human capital	total physical capital [a]	ratio of human capital and physical capital
	deflator for fixed capital formation(1985=100)		
1985	26.98	1.42	19.01
1986	28.05	1.57	17.82
1987	29.98	1.76	17.06
1988	32.74	1.95	16.77
1989	36.84	2.08	17.72

Continue to the next page

1990	37.49	2.20	17.01
1991	37.59	2.37	15.87
1992	37.34	2.61	14.32
1993	37.18	2.94	12.65
1994	45.33	3.34	13.57
1995	52.34	3.80	13.78
1996	61.00	4.29	14.2
1997	69.63	4.79	14.53
1998	74.91	5.36	13.98
1999	81.69	5.92	13.81
2000	88.32	6.54	13.51

*. Use the deflator based on 1952 to convert to the deflator based on 1985 (See Table C.9).

Moreover, we also compare our human capital estimates with the estimated total physical capital stock in China. There are a few estimates of China's capital stock. In Table IV.1.3 the estimated capital stock is estimated by Zhang, Wu and Zhang (2004) published in Economic Research, a leading academic journal in China. In Table IV.1.4, we use the capital stock estimates reported in Holz (2006). In both tables, we use the same deflators reported in the paper to calculate the human capital stock, respectively.

As can be seen in Figure IV.1.4 and Figure IV.1.5, in both cases, the total human capital is much higher than total physical capital. More specifically, human capital is about 10-20 times of the amount of physical capital. This is not surprising, given that in most countries human capital accounts for over 60% of national wealth (which also include natural resources). On the other hand, the ratio of human capital to physical capital appears to be declining continuously, based on both estimates of physical capital. It is unclear whether such a trend indicates that the Chinese government has overly weighted toward physical capital investment relative to human capital investment.[15]

[15] Heckman (2005) and Liu (2007) also find over-investment of physical capital and under-investment of human capital in China during the reform period.

Table IV.1.4 Total human capital and midyear real original value of fixed assets (Holz, 2006), 1985-2003, in trillions

year	total human capital	midyear real original value of fixed assets [a]	ratio of total human capital and fixed assets
1985	26.98	1.73	15.56
1986	28.05	1.95	14.38
1987	29.99	2.18	13.78
1988	32.75	2.43	13.49
1989	36.84	2.70	13.62
1990	37.50	2.97	12.62
1991	37.25	3.26	11.44
1992	36.27	3.58	10.12
1993	35.67	3.94	9.06
1994	43.48	4.32	10.06
1995	50.23	4.75	10.58
1996	58.55	5.24	11.18
1997	66.82	5.78	11.56
1998	71.89	6.35	11.33
1999	78.41	6.94	11.30
2000	84.77	7.56	11.22
2001	90.89	8.19	11.10
2002	96.66	8.87	10.89
2003	103.40	9.66	10.70

*. Scrap value deflated using deflator of earlier period (1985=100) (See Table C.9)

Figure IV.1.4 Total human capital and physical capital (Zhang *et. al.* 2004), 1985-2000

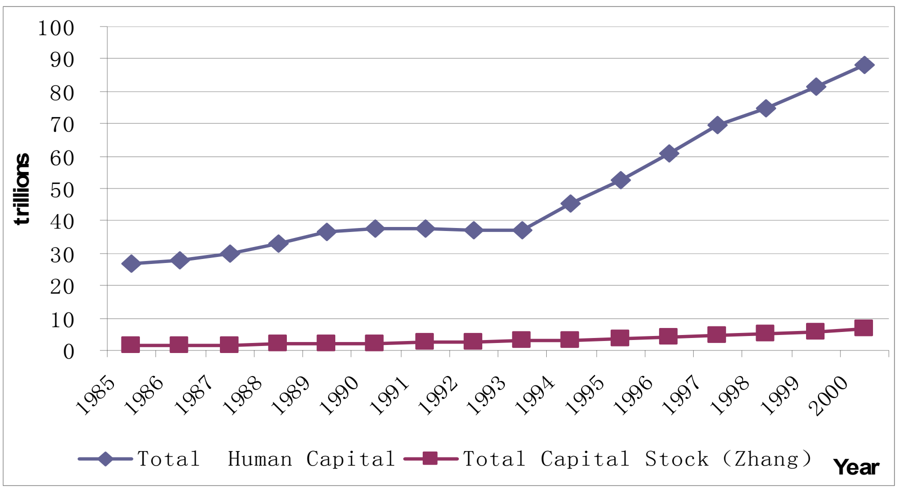
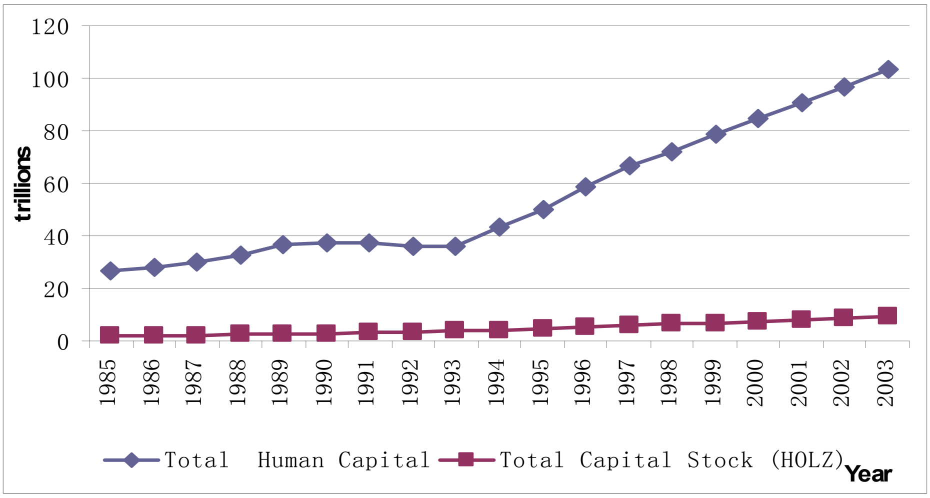


Figure IV.1.5 Total human capital and physical capital (Holz，2006), 1985-2003



IV. 2 The trend of total human capital stock

In order to discuss the trend of the total human capital in China, we use CPI as deflator to calculate the real values. One reason is that other published deflators are not available for later years; and the other reason is that, as can be seen above, the results based on CPI are smaller than that based on capital deflators reported in those two studies. Thus, we give more conservative estimates of human capital in China.

From 1985 to 2007, the total human capital increased from RMB 26.98 trillion to 118.75 trillion, an increase of more than three-fold. The average annual growth for this period is 6.74% per year, considerably lower than economic growth.[16] Over the same period, the Chinese economy grew at an annual rate of 9.33%.[17] This helps explain the declining ratio of human capital to GDP. However, such a growth rate is much higher compared to that in other countries. For example, for

[16] In calculating annual average growth rate in this report, we calculate annual growth rate using the difference of logarithm for every year, and then take average across years.
[17] The data come from "China Statistical Yearbook 2008", Table 2-4.

1970-2000, the annual average growth of human capital in Canada was 1.7% per year (Gu and Wang 2009). Moreover, the growth of human capital accelerated after 1994. The average annual growth for 1985-94 is 5.11%, and for 1995-07 is 7.86%.

The results based on six education categories give similar trend (Figure IV.2.1). From 2000 to 2007, the total human capital increased from RMB 73.5 trillion to 122.38 trillion. The average annual growth rate for this period was 7.28%. The total human capital for male is higher than that for female (Figure IV.2.2). One reason is the earlier retirement age for women (age 55, vs. age 60 for men based on China labor law), and thus men have longer time to generate income in the market. The other reason is higher educational attainment for men. Moreover, the male-female income gap has been on rising. The results based on six education categories shows similar trends.

Figure IV.2.1 Total real human capital by education categories, 1985-2007

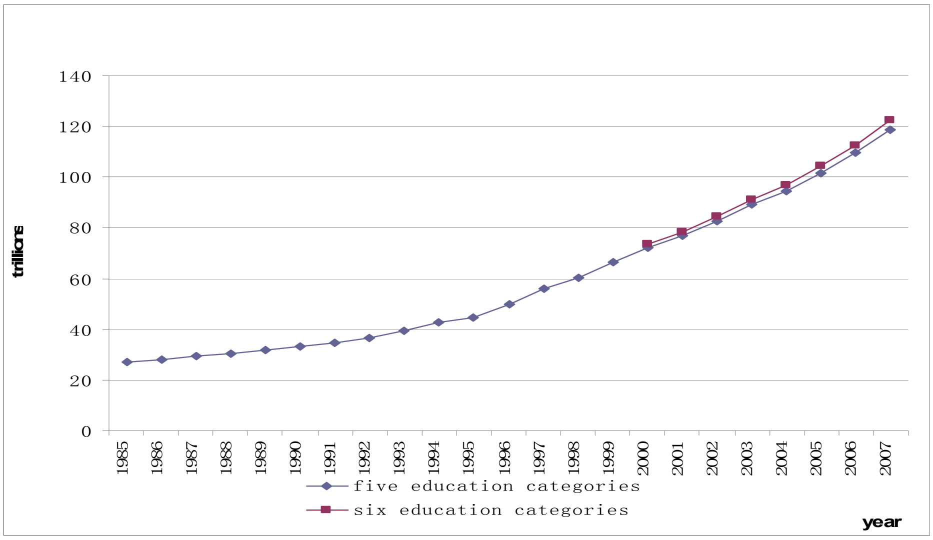


Figure IV.2.2 Total real human capital by gender, 1985-2007

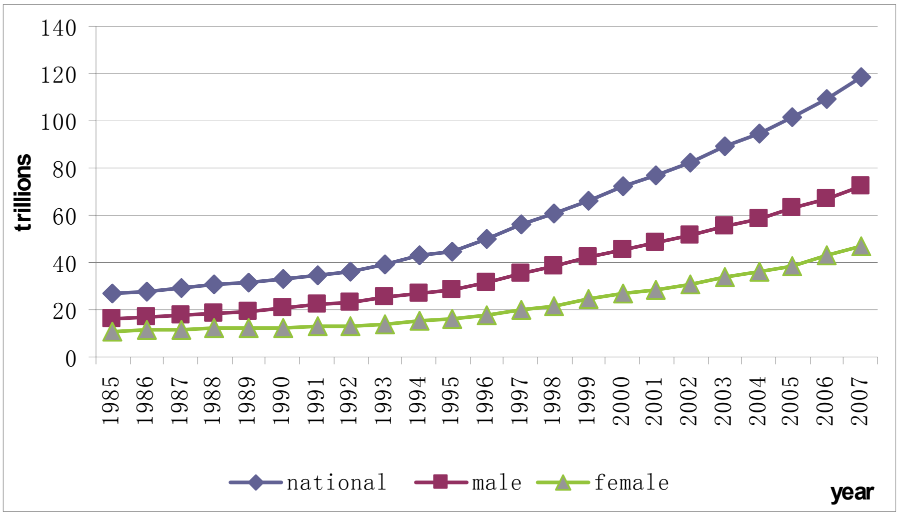


Figure IV.2.3 Total real human capital by urban and rural, 1985-2007

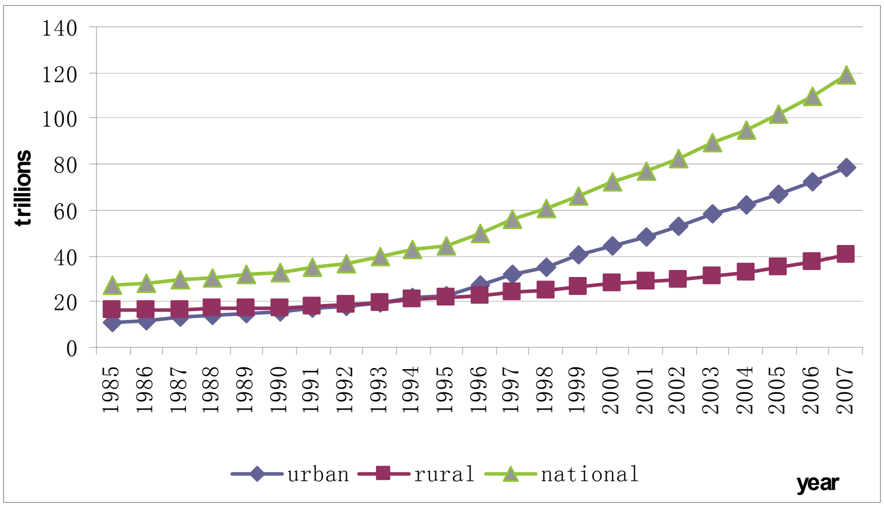
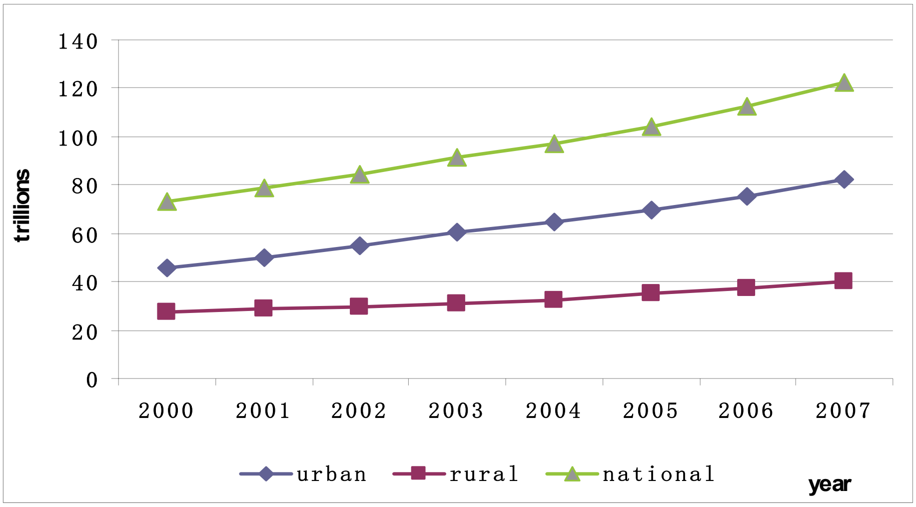


Figure IV.2.3 shows the total human capital for urban and rural China separately. Before 1995, the amount of total human capital in both areas was very close. In fact, rural human capital was even larger than that in the urban area until 1993. Since 1995, however, the human capital in the urban area has been rising much more rapidly. The total human capital for the rural area was 16.03 trillion in 1985 and 40.25 trillion in 2007; and for the urban area it was 10.95 trillion and 78.50 trillion, respectively. In this period, the annual growth rates of human capital were 4.19% (4.99%

after 1995) and 8.95% (9.90% after 1995) for rural and urban areas, respectively. The urban-rural gap in the estimated human capital stock increased from 1.24 trillion in 1995 to 38.25 trillion in 2007, growing at an annual rate of 28.55%. Figure IV.2.4 shows the total human capital estimates in urban and rural areas based on six education categories. The trends are similar to those based on five education categories.[18]

Figure IV.2.4 Total real human capital by urban and rural, 2000-2007



There are several reasons for such a trend. First, in early years, the rural population dominated, and thus had larger amount of human capital. For example, in 1985, there were 733 million people in rural areas, which were more than three times the urban population of 229 million. By 2007, however, the population in rural China reduced to 608 million, much closer to the urban population of 507 million. This change was, to a large extent, a result of the rapid urbanization during the course of economic transition as well as a large scale rural-urban migration.

The second reason is the education gap between the urban and rural population. In urban areas, the population with education at college or

[18] However, our estimates for the rural area are rather conservative because we assume the same male retirement age of 60 and female retirement age of 55 as in the urban area. In fact, many rural residents continue to work after these ages.

above accounted for 2.47% of the total population in 1985. This proportion increased to 13.01% by 2007. While in rural areas, the corresponding figures were 0.074% in 1985 and 0.93% in 2007.

Figure IV.2.5 Total urban human capital by gender, 1985-2007

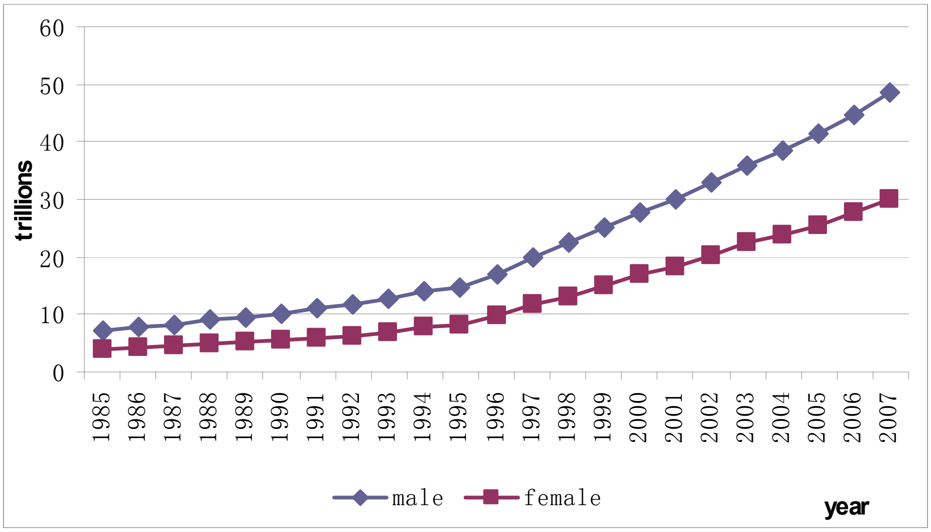
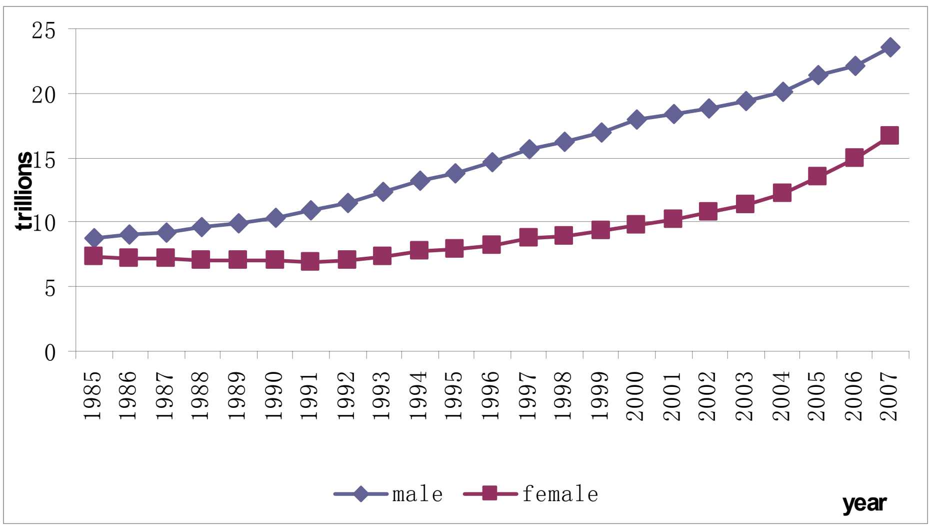


Figure IV.2.6 Total rural human capital by gender, 1985-2007



Figures IV.2.5 and IV.2.6 show the trends of male and female human capital estimates in urban and rural areas, respectively. Male and female human capital estimates in the urban area exhibit similar trend. But the gender gap seems to be widening. The gender-based human capital estimates for the rural population painted a somewhat different picture. In the later part of the period, the growth of human capital of males seems to have slowed down while that of females seems to have sped up, and therefore the gender gap became narrower. This result is probably caused by two factors: i) a disproportionate rural-to-urban migration in favor of men; and ii) an increase in education for women in rural areas. The reduction of gender gap in the rural area is consistent with the rising gender disparity in the urban area. Similar patterns emerge from the results based on six education categories (Figures IV.2.7 and IV.2.8).

Figure IV.2.7 Total urban human capital by gender, 2000-2007

Figure IV.2.8 Total rural human capital by gender, 2000-2007

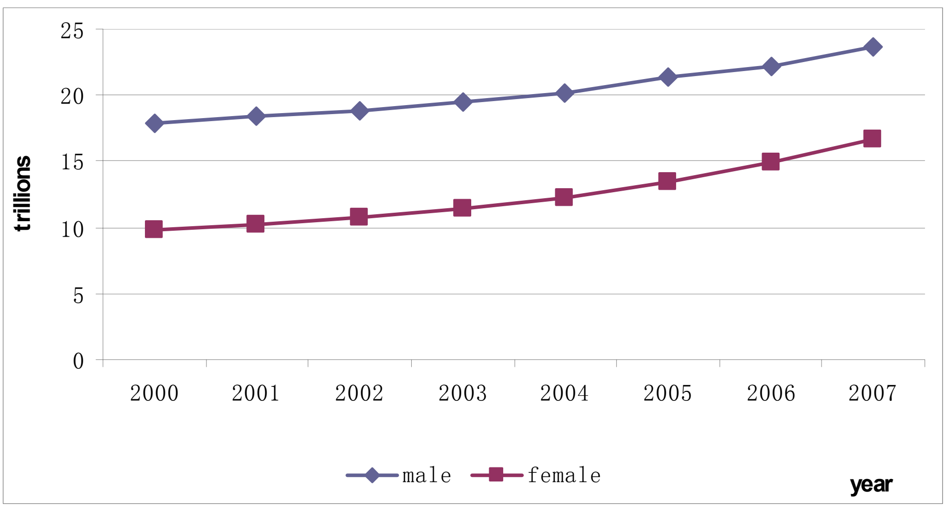


Table IV.2.1 Total human capital index, 1985-2007 (1985=100)

Year	total human capital	male total human capital	female total human capital	urban total human capital	rural total human capital
1985	100	100	100	100	100
1986	104	105	102	108	101
1987	109	111	107	118	103
1988	113	118	108	126	105
1989	117	123	110	134	106
1990	122	129	112	143	109
1991	128	138	114	153	111
1992	135	146	120	164	115
1993	146	159	128	181	123
1994	158	171	140	198	131
1995	165	179	145	209	135
1996	184	200	162	245	143
1997	208	225	183	289	152
1998	224	243	197	322	157
1999	246	266	219	367	164
2000	268	288	239	406	173
2001	286	306	256	442	179
2002	306	326	279	484	184
2003	331	348	305	533	192
2004	351	370	324	568	202
2005	377	397	349	611	217
2006	406	421	384	661	232
2007	440	454	420	717	251

Finally we calculate human capital index using 1985 as the base year and set its value at 100. The results for each group are reported in Table IV.2.1. Figure IV.2.9 shows the index of total human capital, and Figures IV.2.10 and IV.2.11 show the index by gender for urban and rural areas, respectively.

Figure IV.2.9 The index of total human capital, 1985-2007

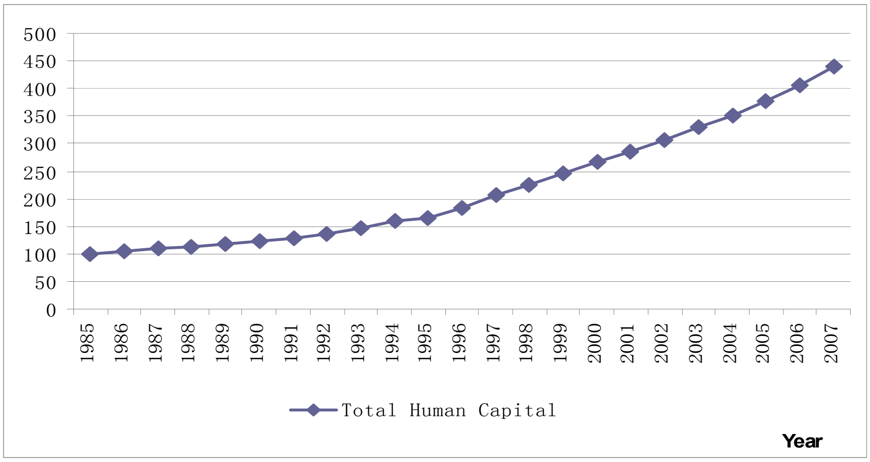


Figure IV.2.10 The index of total human capital by gender, 1985-2007

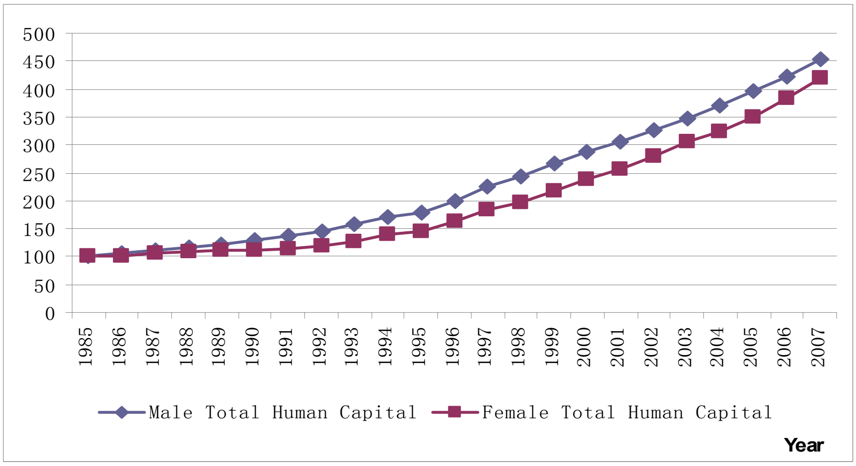
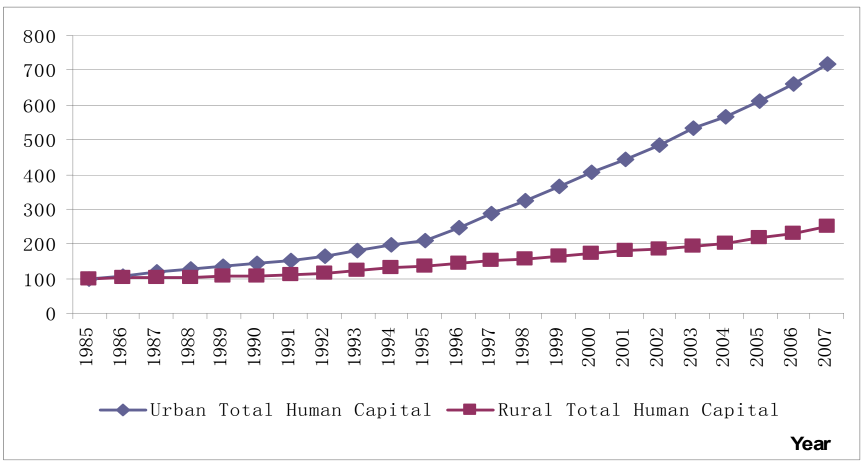


Figure IV.2.11 The index of total human capital by urban and rural, 1985-2007



IV.3 Per capita human capital

The increase in the total human capital can be caused by population growth, demographic changes (e.g., the size of retirement group), rural-urban migration or urbanization (e.g., an individual can achieve higher value of human capital by moving from rural to urban area), higher educational attainment, higher rates of return to education, higher rates of return to on-the-job training, etc. In order to get further information on the dynamics of human capital in China, we calculate per capita human capital, i.e., the ratio of total human capital over non-retired population (Table IV.3.1).

Figures IV.3.1 and IV.3.2 show per capita human capital based on 5- and 6-education categories, respectively. Based on 5-education category, the per capita human capital was RMB 28,044 in 1985, RMB 41,500 in 1995, and RMB 106,462 in 2007. From 1985 to 2007, per capita human capital increased 2.80 times; while over the same period, per capita real GDP increased 6.68 times, much faster than the growth of per capita human capital. Per capita human capital has been increasing since 1985, and the growth accelerated from 1995. The average annual growth rate was 3.9% from 1985 to 1994, and 7.5% from 1995 to 2007. The growth rate in the later period is almost twice as high as that in the earlier period.

Table IV.3.1 Real per capita human capital and real per capita GDP (1985 yuan)

year	real per capita human capital			real per capita GDP
	national	urban	rural	
1985	28,044	47,874	21,856	858
1986	28,755	49,445	22,018	934
1987	29,717	51,671	22,269	1,042
1988	30,473	53,269	22,517	1,160
1989	31,081	54,687	22,655	1,207
1990	31,933	56,851	22,921	1,253
1991	33,170	59,528	23,409	1,368
1992	34,622	62,253	24,160	1,563

Continue to the next page

1993	37,201	66,830	25,728	1,781
1994	39,996	71,541	27,499	2,014
1995	41,500	73,996	28,340	2,234
1996	45,804	81,441	30,256	2,458
1997	51,063	90,412	32,607	2,686
1998	54,672	95,361	34,199	2,897
1999	59,638	102,885	36,332	3,117
2000	64,355	108,553	38,896	3,380
2001	68,627	113,484	41,135	3,661
2002	73,503	119,520	43,461	3,993
2003	79,330	126,543	46,493	4,394
2004	84,281	131,048	50,040	4,837
2005	91,147	137,882	55,208	5,341
2006	98,080	146,019	59,796	5,964
2007	106,462	154,803	66,164	6,675

These growth rates are very high compared to those for Canada and the United States. Per capita human capital for Canada basically remained constant during 1980-2000 and even declined at an annual rate of -0.2% during 2000-2007 (Wu and Ambrose 2009). Per capita human capital in the United States also basically remained constant during 1994-2006 (Christian 2009). Such a huge difference is probably caused by the dramatic economic growth since 1978, rapid expansion of education, transition toward market-oriented system (so that human capital can realize much higher value), and rural-urban migration.

Figure IV.3.1 Real per capita human capital by gender, 1985-2007

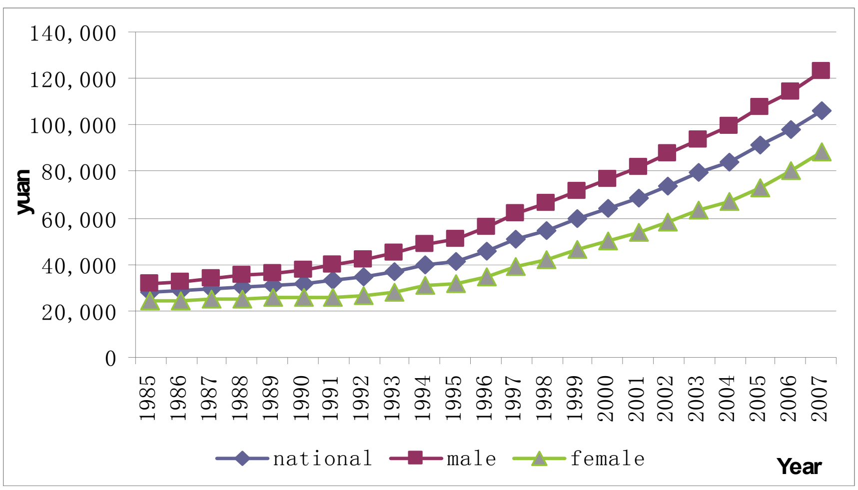
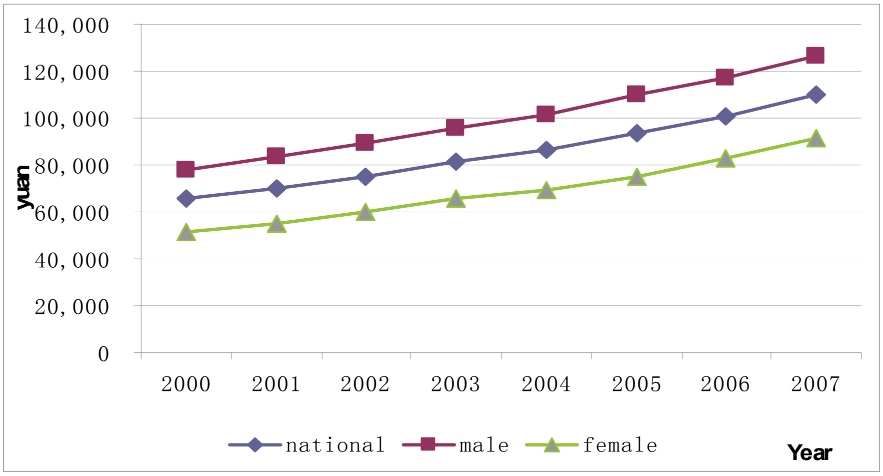


Figure IV.3.2 Real per capita human capital by gender, 2000-2007



Per capita human capital shows a similar trend for males and females. Specifically, the average annual growth rate for 1985-1994 was 4.8% for males and 2.6% for females; the average annual growth rate for 1995-2007 was 7.2% for males and 8.1% for females.

Clearly, the percentage point increase in the growth rates between the two periods is substantially greater for females than for males. In fact, from 1996 onward, the growth rate was lower for males than for females.

Figures IV.3.3 and IV.3.4 show per capita human capital for urban and rural areas based on two alternative classifications of education. Based on 5-education category, in 1985, per capita human capital is 47,874 in the urban area and 21,856 in the rural area; the corresponding numbers become 154,803 and 66,164, respectively, in 2007. The absolute size of the urban-rural gap has been on the rise. The annual growth rate was 5.33% for the urban area (4.46% for 1985-1994 and 5.94% for 1995-2007), and 5.03% for the rural area (2.55% for 1985-1994 and 6.75% for 1995-2007). Therefore, the urban-rural gap was widening for 1985-1994, while it has narrowed thereafter. The wide urban-rural gap raises concern for the increasing disparity between these two areas. Based on Fleisher, Li and Zhao (2009), human capital is a significant contributing factor to economic growth (total factor productivity). Therefore, such a trend in human capital can worsen the urban-rural inequality in China.

Figure IV.3.3 Real per capita human capital by urban and rural, 1985-2007

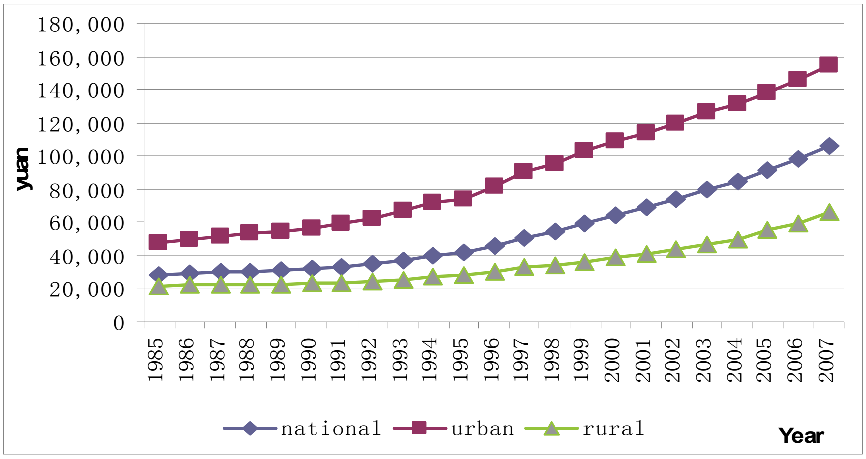
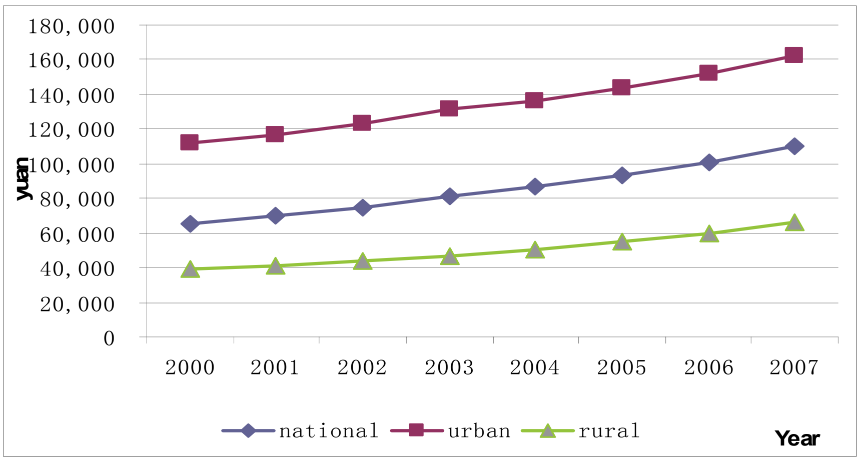


Figure IV.3.4 Real per capita human capital by urban and rural, 2000-2007



Figures IV.3.5 and IV.3.6 show the gender differences for urban and rural areas, respectively. The patterns are similar to that of total human capital. In particular, per capita human capital for males and females show similar trend in the urban area, but per capita human capital grew faster for females than males in the rural area in recent years. From 1985 to 2002, rural male per capita human capital grew at an annual rate of 4.90% compared to 2.78% for females; from 2003 to 2007, however, the growth rates were 6.72% and 11.06%, respectively. Although both male and female growth rates have increased, the female growth rate has increased much more than the male.

Figure IV.3.5 Urban real per capita human capital, 1985-2007

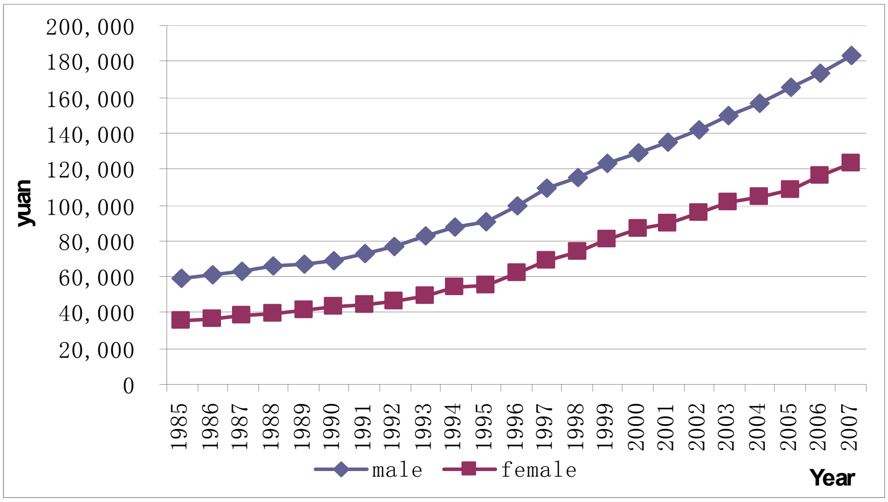
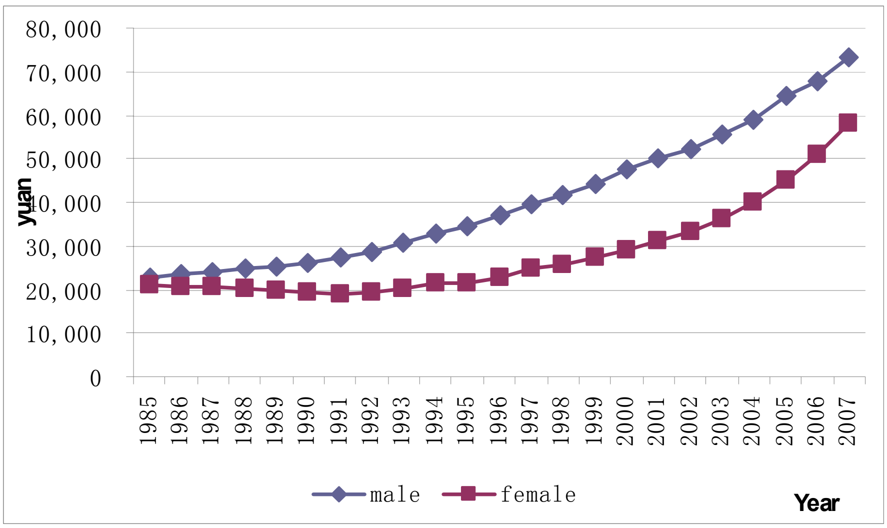


Figure IV.3.6 Rural real per capita human capital, 1985-2007



We also construct per capita human capital index with its corresponding value in 1985 set as 100 (Table IV.3.2). Figures IV.3.7and IV.3.8 show various per capita human capital indexes.

Table IV.3.2 Per capita human capital index 1985-2007 (1985=100)

year	average human capital	male average human capital	female average human capital	urban average human capital	rural average human capital
1985	100	100	100	100	100
1986	103	104	101	103	101
1987	106	108	103	108	102
1988	109	112	103	111	103
1989	111	115	105	114	104
1990	114	120	105	119	105
1991	118	127	106	124	107
1992	123	133	110	130	111
1993	133	144	116	140	118
1994	143	154	126	149	126
1995	148	161	130	155	130
1996	163	177	143	170	138
1997	182	197	161	189	149
1998	195	211	172	199	156
1999	213	228	190	215	166
2000	229	245	207	227	178
2001	245	261	221	237	188
2002	262	278	239	250	199
2003	283	298	261	264	213
2004	301	317	277	274	229
2005	325	343	300	288	253
2006	350	364	330	305	274
2007	380	392	363	323	303

Figure IV.3.7 Real per capita human capital index by gender, 1985-2007

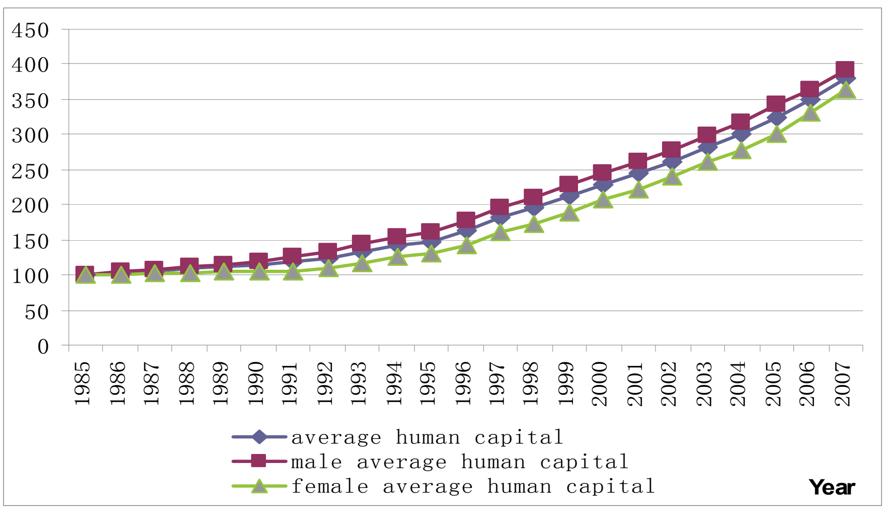
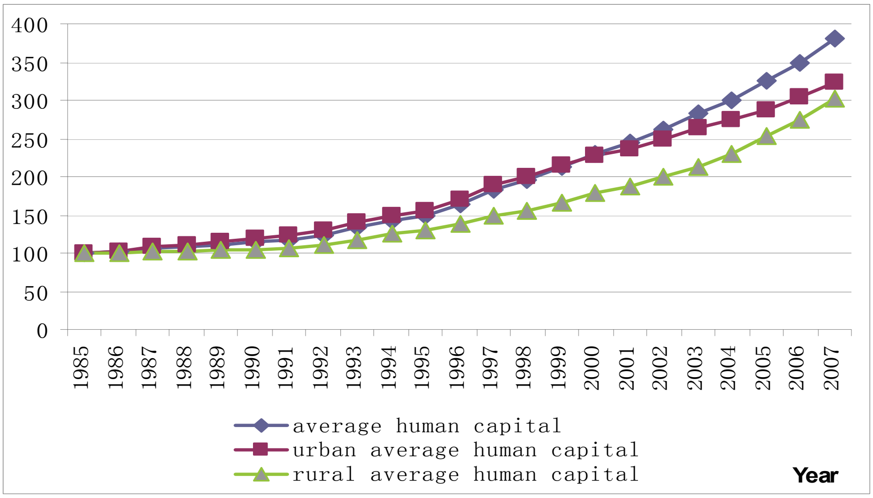


Figure IV.3.8 Real per capita human capital index by urban and rural, 1985-2007



IV.4 Divisia indexes

Two partial alternative indexes are constructed for real human capital. The first aggregates by gender and the second over five education levels. These indexes are partial Divisia indexes (Gu and Wong, 2009) as they do not separately identify all of the components of human capital: gender, age, education, and location and they are first order indexes. Nonetheless these indexes are of interest because they show the differential trends in human capital by gender compared to education. These indexes are shown in Table IV.4.1 and Figures IV.4.1~2.

The education index is constructed as follows. The growth rate of aggregate human capital stock is calculated as a weighted sum of the growth rates of the number of individuals across different educational categories:

$$d\ln K^{e} = \sum_{e} \bar{v}_{e} d\ln L_{e}$$

where $dlnK^e$ denotes the growth rate of aggregate human capital and L_e denotes the number of individuals with education level e. Also,

$$d \ln L_e = \ln L_e(y) - \ln L_e(y-1)$$

where y denotes the year. The weights are given by nominal human capital shares for each educational level:

$$\bar{v}_e = \frac{1}{2}\left[v_e(y) + v_e(y-1)\right] \qquad v_e = \frac{Mi_e}{\sum_e Mi_e}$$

where Mi_e is the nominal human capital of individuals with education level e.

The partial index for gender is estimated in a similar fashion, with the subcomponents being male and female instead of education categories. The rate of growth of the education index is substantially higher than that for the gender index. Given the substantial increase in educational attainment over this time period, this is not surprising. From 1986 to 1994, the gender index grew at a 1.15% rate compared to a 14.09% rate for the education index. From 1994 to 2007, the corresponding numbers are 0.33% and 5.5%, respectively.

Table IV.4.1 Partial Divisia index for gender and education 1986-2007 (Base year: 2001, in trillions)

year	gender	five education levels
1986	228.78	51.91
1987	231.95	53.35
1988	235.82	80.16
1989	239.48	103.28
1990	243.02	123.51
1991	245.40	134.18
1992	247.38	143.32
1993	249.10	153.21
1994	250.75	160.21
1995	252.06	167.12
1996	254.81	183.31
1997	257.44	199.12

Continue to the next page

1998	259.82	215.72
1999	261.83	232.16
2000	263.73	249.00
2001	263.75	263.75
2002	263.89	274.01
2003	263.84	281.84
2004	263.27	289.73
2005	261.87	300.10
2006	261.82	314.48
2007	261.80	326.40

Figure IV.4.1 Partial Divisia index for gender, 1986-2007

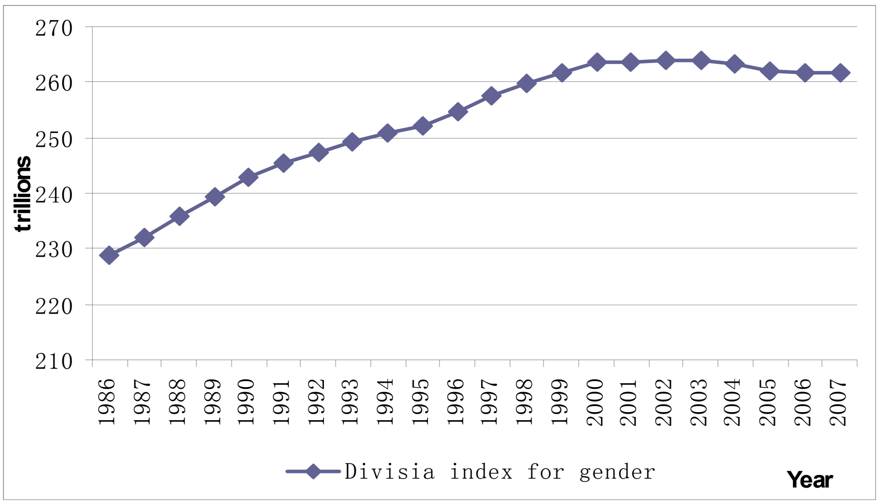
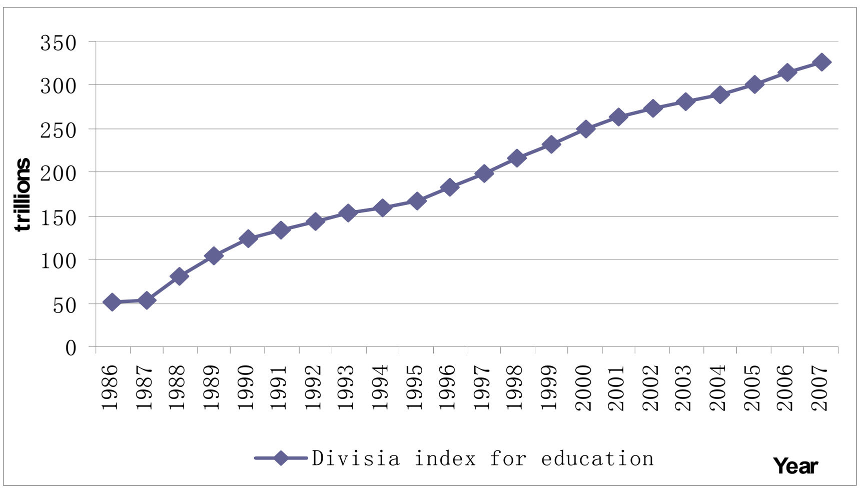


Figure IV.4.2 Partial Divisia index for education, 1986-2007

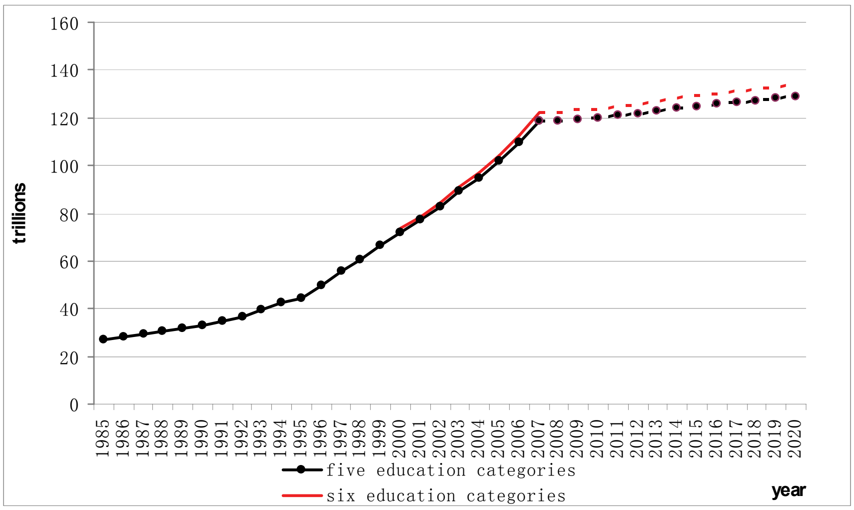


IV.5 Human capital in China 2008-2020: a projection

In order to understand future trend of human capital in China, we estimate human capital for 2008-2020. In particular, we forecast population in different age, gender and education groups using the perpetual inventory method, and then estimate human capital using the Jorgenson-Fraumeni method. For simplicity, we keep all other related data and parameters at their 2007 values.[19]

If we only project population in different age, gender and education groups for 2008 to 2020 while keeping other variables at their 2007 values, the change in human capital will mainly reflect the change in population composition. Figure IV.5.1 shows that results based on 5- and 6-education categories.

Figure IV.5.1 Total real human capital by education categories, 1985-2020



In both cases, the total human capital increases but at a much slower rate compared to that before 2008. The average annual growth rate is 0.61%, based on 5-education-category. This is much lower than the

[19] Due to data limitation, we use the average values of year 1995 and 2000 for age, gender and education based employment rates.

average annual growth of 6.74% for 1985-2007. There are several reasons for the slower growth. First, the return to education is kept at 2007 level, but was rising before that period. Return to education has a strong effect on lifetime earnings. Second, population growth will slow down in China due to the one-child policy. Third, it is expected that the growth of human capital will slow down when the economy gets closer to its steady state, including wage growth, returns to schooling, etc.

A similar pattern can be seen in male and female total human capital and per capita human capital (Figures IV.5.2 and IV.5.3). Interestingly the trends are quite different for urban and rural areas. As Figure IV.5.4 shows, urban human capital continues to increase throughout the entire period. However, the rural human capital declines. This is probably caused by the continuing declining of rural population, as a result of urbanization and rural-urban migration. However, the per capita human capital (Figure IV.5.5) in the rural area is quite flat and does not show a downward trend.

Figure IV.5.2 Real total human capital by gender, 1985-2020

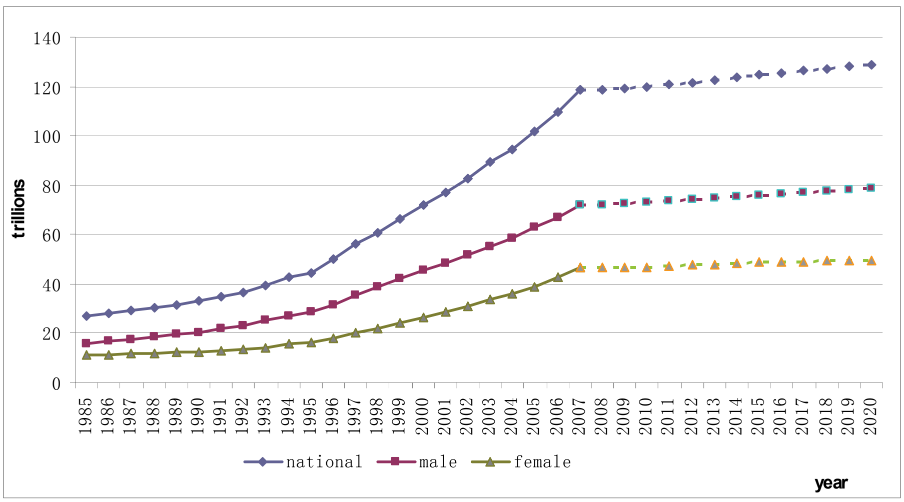


Figure IV.5.3 Real per capita human capital by gender, 1985-2020

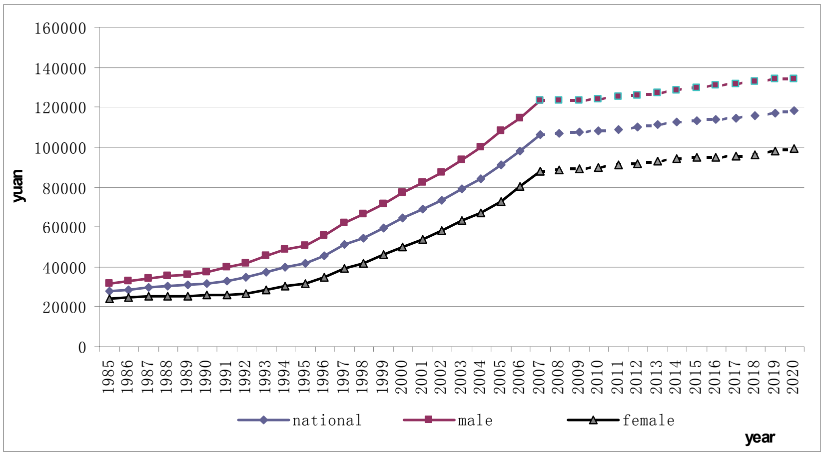


Figure IV.5.4 Real total human capital: urban, rural and national, 1985-2020

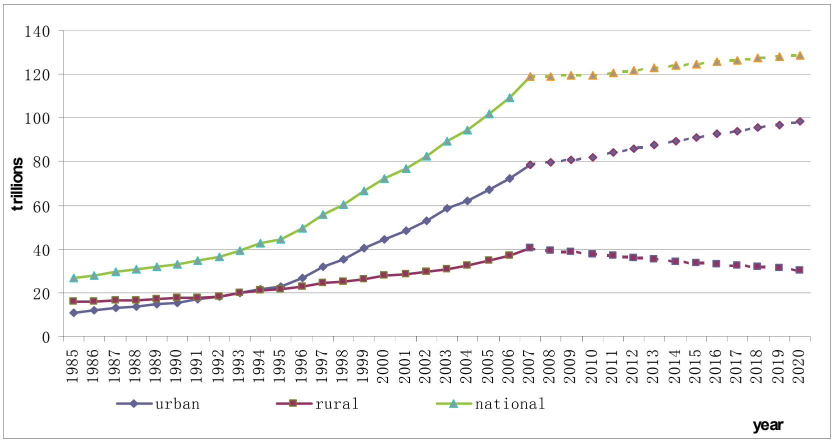
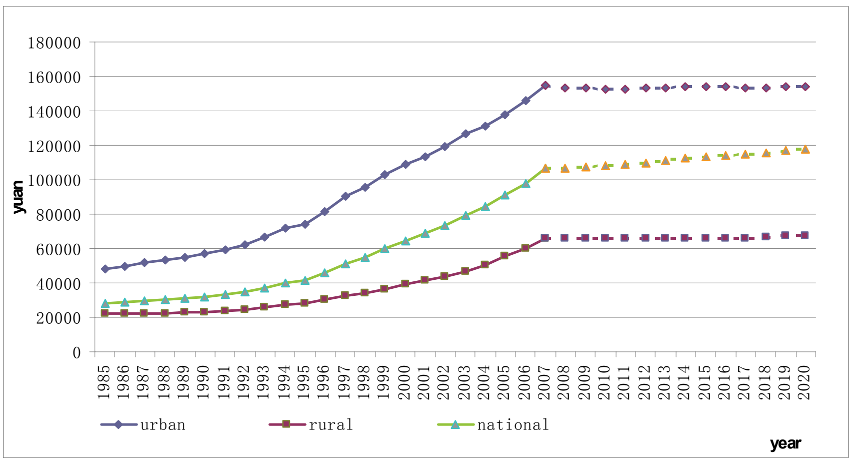


Figure IV.5.5 Per capita real human capital: urban, rural and national, 1985-2020



V Conclusions

In this report, we presented our estimates of China's human capital for 1985-2007, using J-F lifetime income approach. We calculated total human capital at the national level, for urban and rural, and for male and female, as well as per capita human capital. We also constructed various human capital indexes, including partial Divisia quantity indexes. We projected the trend of human capital in one scenario for up to year 2020.

Our main findings are summarized below：

First, for the period of 1985-2007, China's total human capital increased more than three times, with an annual growth rate of 6.74%. This growth rate is much higher compared to other countries. Moreover, the growth of human capital accelerated after 1994, and the average annual growth for 1995-07 is 7.86%.

Second, the total human capital in urban area increased at a much higher rate than in rural area over the period 1985-07. The annual average growth rates are 8.95% and 4.19% respectively for urban and rural areas. The total human capital in urban area surpassed that in rural area in 1993. The urban-rural gap has been widening rapidly, probably because of urbanization, large-scale rural-urban migration, and increase in educational attainment.

Third, per capita human capital also increased rapidly from 1985-2007, with a higher growth rate since 1995. Interestingly, before 1995 total human capital increased faster than per capita human capital on average, while since 1995, both have grown at a similar average annual rate. This result indicates that in recent years, the growth of human capital is mostly driven by factors such as increases in educational attainment, not by population growth.

Fourth, the gender gap in total human capital has been widening at the national level. However, the gender difference in per capita human capital appears to be narrowing down.

Fifth, the partially education-based human capital index grew at a much higher rate than the gender-based index. This indicates the greater impact of education on China's human capital accumulation.

On the other hand, our results also show that, compared to GDP and physical capital, human capital grew at a slower pace. More specifically, the ratio of human capital to GDP decreased from approximately 30 in 1985 to 18 in 2007; and the ratio of human capital to physical capital also declined from 16-19 in 1985 to 11~12 in 2003, these findings indicates that the Chinese government should invest more in human capital, especially compared to physical capital investment.

The gap in total human capital and per capita human capital between urban and rural areas has been increasing. Thus, in order to reduce urban-rural inequality, more investment in human capital should be directed to the rural area.

Finally, our projection to 2020 shows that, if we keep everything else at the 2007 level and only allow population to change, the growth of total human capital and per capita human capital will slow down after 2007. The amount of total human capital will even decline in rural China. Therefore, more active policies on human capital investment should be adopted in order to maintain the high speed growth.

Our future work includes: i) finding more data to improve estimates of lifetime earnings and other related variables; ii) refining the estimation of some related parameters and data; and iii) refining our projections of future incomes and testing the effects of various policy scenarios on human capital accumulation.

Reference List

Abraham, Katharine (2005), *Beyond the Market: Designing Nonmarket Accounts for the United States*, National Academies Press, Washington, D.C.

Ahlroth, Sofia, A. and Bjorklund, A. Forslund (1997), "The Output of the Swedish Education Sector", *Review of Income and Wealth* Volume 43, Number 1, pp.89-104.

Ashenfelter, Orley and Krueger, Alan (1994), "Estimates of the Economic Return to Schooling from a New Sample of Twins", *American Economic Review* 84, December, pp.1157-73.

Becker, G. (1964), *Human Capital*, 2nd edition, Columbia University Press, New York.

Bureau of Statistics of China (2008), China Population Statistical Yearbook 2008, China Statistics Press, Beijing.

Cai Fang and Wang,Dewen(1999), "The Sustainability of China's Economic Growth and Labor Contributions", Economic Research (Chinese) 10, pp.62-68

Christian, Michael S. (2009), "Human Capital Accounting in the United States: 1994 to 2006", paper presented at the Canadian Economic Association Annual Conference, May 29.

Coleman, J. (1990), Foundations of Social Theory, Belknap, 304

Démurger, Sylvie (2001), "Infrastructure Development and Economic Growth: An Explanation for Regional Disparities in China?" Journal of Comparative Economics 19, pp.95-117.

Ederer, Peer (2006). "Innovation at Work: The European Human Capital Index", The Lisbon Council Policy Brief, in conjunction with Deutschland Denken! and Zeppelin University, Brussels, October 12

Ederer, Peer, Philipp Schuller, and Stepham Willms (2007), "Innovation at work: The European Human Capital Index", The Lisbon Council Policy Brief, Volume 2, Number 3, Brussels

Fleisher, Belton., and Chen, Jian (1997), "The Coast-Noncoast Income Gap, Productivity and Regional Economic Policy in China", Journal of Comparative Economics 252: pp.220-236.

Fleisher, Belton., Sabirianova, Klara., and Wang, Xiaojun (2005), "Returns to Skills and the Speed of Reforms: Evidence from Central and Eastern Europe, China, and Russia.", Journal of Comparative Economics 33, issue 2, pp.351-370.

Fleisher, Belton and Wang, Xiaojun (2004), "Skill Differentials, Return to Schooling, and Market Segmentation in a Transition Economy: the Case of Mainland China", Journal of Development Economics 73, pp.315-328.

Fleisher, Belton, Li Haizheng and Zhao Minqiang (2009), "Human Capital, Economic Growth, and Regional Inequality in China", *Journal of Development Economics*, forthcoming.

Fraumeni, Barbara M. (2006), "A Human Capital Approach to Measuring Government Education Output", presentation at the OECD/ONS/Government of Norway Workshop on the Measurement of Non-market Output and Health, London, England, UK, October 4.

Fraumeni, Barbara M. (2008a), "Human Capital and Investment in Education: A Streamlined Approach", presentation at the Fondazione Giovanni Agnelli/OECD Workshop on the Measurement of Human Capital, Turin, Italy, November 3.

Fraumeni, Barbara M. (2008b), "Human Capital: From Indicators and Indexes to Accounts," paper presented at the Fondazione Giovanni Agnelli/OECD Workshop on the Measurement of Human Capital, Turin, Italy, November 4.

Fraumeni, Barbara M. (2009), "Further Comments: Results by Country and Future Efforts" presentation at a China Center for Human Capital and Labor Market Research seminar at the Central University of Finance and Economics, Beijing, China, June 19.

Greaker, Mads and Liu, Gang (2008) "Measuring the Stock of Human Capital for Norway: A Lifetime Labour Income Approach," paper presented at the Fondazione Giovanni Agnelli/OECD Workshop on the Measurement of Human Capital, Turin, Italy, November 3.

Griliches, Zvi, "Estimating the Returns to Schooling." *Econometrica* 45, 1977, pp.1-22.

Gu, Wulong and Wong, Ambrose (2008), "Human Development and its Contribution to the Wealth Accounts in Canada," paper presented at the Fondazione Giovanni Agnelli/OECD Workshop on the Measurement of Human Capital, Turin, Italy, November 3.

Gu, Wulong and Wong Ambrose (2009), "Human Development and its Contribution to the Wealth Accounts in Canada," paper presented at the Canadian Economic Association Annual Conference, May 29.

Heckman, James J. (2005), “China's Human Capital Investment”, China Economic Review 16, pp.50-70.

Holz, Carsten A. (2006), “New Capital Estimates for China”, China Economic Review 17, pp.142-185.

Hou,Yafei and Cao Yin(2000), ”Analysis of the Quality of Human Capital Stock”, Chinese Journal of Population Science(Chinese) 6, pp.43-48

Hu,Angang(2002), ”From the Most Populous Country to A Country with Great Power of Human Capital:1980-2000”, Chinese Journal of Population Science(Chinese) 5, pp.1-10

Hu,Yongyuan(2005), ”Human Capital and Economic Growth: A Co-integration Analysis”, Science and Technology Management Research(Chinese) 4, pp.88-90

Jorgenson, Dale W. and Fraumeni, Barbara M. (1989), “The Accumulation of Human and Non-Human Capital, 1948-1984,” in R. Lipsey and H. Tice eds., *The Measurement of Saving, Investment and Wealth*, Chicago, University of Chicago Press, NBER, pp. 227-282.

Jorgenson, Dale W. and Fraumeni, Barbara M. (1992a), “Investment in Education and U.S. Economic Growth,” *Scandinavian Journal of Economics,* Vol. 94, supplement, pp. S51-70.

Jorgenson, Dale W. and Fraumeni, Barbara M. (1992b), “The Output of the Education Sector,” in Z. Griliches, T. Breshnahan, M. Manser, and E. Berndt eds., *The Output of the Service Sector*, Chicago, NBER, 1992, pp. 303-341.

Keeley, Brian (2007), *Human Capital, How What You Know Shapes Your Life*, OECD Insights, Paris.

Kendrick, J. (1976), The Formation and Stocks of Total Capital, NBER, Columbia University Press, New York, N.Y.

Koman, R., and Marin, D. (1997), “Human Capital and Macroeconomic Growth: Austria and Germany 1960-1997. An Update.” IAS Economics Series No. 69.

Laroche, M. and Merette, M. (2000), “Measuring Human Capital in Canada.” Ministry of Finance of Canada.

Le, Trinh Van Thi, Gibson, John, and Oxley, Les (2005), “Measuring the Stock of Human Capital in New Zealand,” *Mathematics and Computers in Simulation*, Volume 68, Issue 5-6, May, pp. 485-98.

Le, Trinh, Gibson, John, and Oxley, Les (2003), “Cost- and Income-Based Measures of Human Capital.” *Journal of Economic Surveys* 17, 3: pp.271-307.

Le, Trinh, Gibson, John, and Oxley, Les (2005), "Measures of Human Capital: A Review of the Literature." New Zealand Treasury Working Paper 05/10.

Li, Haizheng (2003), "Economic Transition and Returns to Education in China." *Economics of Education Review* 2 317-328.

Liu, Zhiqiang (1998), "Earnings, Education, and Economic Reforms in Urban China." *Economic Development and Cultural Change* 46, pp.697-725.

Liu, Zhiqiang (2007), "The External Returns to Education: Evidence from Chinese Cities", Journal of Urban Economics 61 (3), pp.542-564.

Maurer-Fazio, Maggie (1999), "Earnings and Education in China's Transition to a Market Economy: Survey Evidence from 1989 and 1992." China Economic Review 10, pp.17-40.

Mincer, Jacob (1974), Schooling, Experience and Earnings. Columbia University Press. New York.

Mulligan, C. B., and Sala-i-Martin, X. (1997), "A Labor Income-based Measure of the Value of Human Capital: An Application to the States of the United States". *Japan and the World Economy* 9, 2: pp.159-191.

OECD (2001), *The Well-being of Nations: The Role of Human and Social Capital*, OECD, Paris.

Qian,Xuya, and Liu,Jie(2007), "Empirical Study of Human Capital in China", Statistic Research(Chinese) 3, pp.39-45

Schultz, T. (1961), "Investment in Human Capital", *American Economic Review* 51, 1: pp.1-17.

Stroombergen, A., D. Rose and Nana, G. (2002), "Review of the Statistical Measurement of Human Capital", Statistics New Zealand working paper.

Wang,Dejin and Xiang,Rongmei(2006), "Estimates of Human Capital Stock in China",Statistics and Decision (Chinese) 5, pp.100-102

Wang, Xiaojun, Fleisher, Belton, Li, Haizheng, and Li, Shi (2009) "Access to Higher Education and Inequality: A Chinese Experiment." IZA Discussion Paper No. 2823.

Wei, Hui (2008) "Developments in the Estimation of the Value of Human Capital in Australia," paper presented at the Fondazione Giovanni Agnelli/OECD Workshop on the Measurement of Human Capital, Turin, Italy, November 3.

World Bank (1997), "Expanding the Measure of Wealth: Indicators of Environmentally Sustainable Development," Environmentally Sustainable Development Studies and Monographs Series No. 17, Washington, D.C.

World Bank (2006) *Where is the Wealth of Nations, Measuring Capital for the 21st Century*, The International Bank for Reconstruction and Development/The World Bank, Washington, DC.

Yang, Dennis (2005), "Determinants of Schooling Returns during Transition: Evidence from Chinese Cities." *Journal of Comparative Economics* 33, pp.244-264.

Yu,Shujing(2008), "Comprehensive Evaluation and Dynamic Analysis on China's Provincial-level Regional Human Capital", Modern Management Science(Chinese) 4, pp. 36-37

Zhang Fan(2000), "Estimates of Physical Capital and Human Capital in China", Economic Research (Chinese) 8, pp.66-71

Zhang Jun, Wu, Guiying, and Zhang Jipeng (2004), "Compilation of China's Provincial Capital Stock Series Using Perpetual Inventory Method, 1952-2000", Economic Research (Chinese), October.

Zhang, Junsen, Zhao, Yaohui, Park, Alberb, and Song, Xiaoqing (2005), "Economic Returns to Schooling in Urban China, 1988-2001." Journal of Comparative Economics 33, pp.730-752.

Zhou,Delu,(2005), "Population-based Indicators of Human Capital Accounting Theory and Empirical Study", Chinese Journal of Population Science(Chinese) 3, pp. 56-62

Zhou Ya(2004), "Study on the Distribution Differences of China's Human Capital", Education & Economics(Chinese) 2, pp.17-20

Zhu,Pingfang, and Xu,Dafeng(2007), "Estimation of Human Capital in Chinese Cities", Economic Research (Chinese) 8, pp.84-95

Appendix A Population imputation

1. Data collection

1.1 Macro-data

When estimating population by age, gender and education in urban and rural areas, we use the following data sources:

Data	Sources	Notes
National, urban and rural "population aged 6 and over by age, gender and education attainment": 1982,1987, 1990,1995, 2000,2005	• 1982,*China Demographic Statistics Yearbook 1988* edited by Department of Demographic Statistics of National Bureau of Statistics • 1987,*China 1987 1% Demographic Sampling Survey* edited by Department of Demographic Statistics of National Bureau of Statistics • 1990,*China 1990 Census* edited by Census Office of State Council, and Department of Demographic Statistics of National Bureau of Statistics • 1995,*China Demographic Statistics Yearbook.1998* edited by Department of Demographic and Employment Statistics of National Bureau of Statistics • 2000, http://www.stats.gov.cn/tjsj/ndsj/renkoupucha /2000pucha/pucha.htm • 2005, http://www.stats.gov.cn/tjsj/ndsj/renkou/2005 /renkou.htm	
National, urban and rural population aged 0-5 by age and sex: 1982,1987, 1990,1995, 2000,2005	• 1982,*China 1982 Census* edited by State Department Census Office, Department of Demographic Statistics of National Bureau of Statistics • 1987,*China Demographic Statistics Yearbook.1989* edited by Department of Demographic Statistics of National Bureau of Statistics • 1990,*China 1990 Census* edited by State Department Census Office, Department of Demographic Statistics of National Bureau of Statistics • 1995,*China Demographic Statistics Yearbook.1996* edited by Department of Demographic and Employment Statistics of National Bureau of Statistics • 2000, http://www.stats.gov.cn/tjsj/ndsj/renkoupucha /2000pucha /pucha.htm • 2005, http://www.stats.gov.cn/tjsj/ndsj/renkou/2005 /renkou.htm	We assume that those aged 0-5 receive no schooling

Continue to the next page

National, urban and rural population by age and sex: 1982-2007	• *China Demographic Statistics Yearbook.1988—1993* edited by Department of Demographic Statistics of National Bureau of Statistics • *China Demographic Statistics Yearbook.1994—1998,2006* edited by Department of Demographic and Employment Statistics of National Bureau of Statistics • *China Demographic Statistics Yearbook.1999—2005* edited by Department of Demographic and Social Science Statistics of National Bureau of Statistics • *China Demographic and Employment Statistics Yearbook 2007-2008* edited by Department of Demographic and Employment Statistics of National Bureau of Statistics	
Mortality rate by age and sex: 1986,1989-1990, 1994-2007	• *China Demographic Statistics Yearbook：1988—2007*	In the yearbooks of 1988 and 1989, the only mortality rate is of 1986. In the yearbooks of 1992 and 1993, the mortality rate is not separated by age and sex.
Enrollment by education level: 1980-2007	• *Educational Statistics yearbook of China.1987* edited by the Plan and Finance Bureau of National Educational Committee • *Educational Statistics yearbook of China.1989-1992* edited by the Plan and Development Department of National Educational Committee • *Educational Statistics yearbook of China 1993-1996* edited by the Plan and Development Department of National Educational Committee • *Educational Statistics yearbook of China 1997* edited by the Plan and Development Department of National Educational Ministry • *Educational Statistics yearbook of China.1998-2007* edited by the Development and Plan Department of National Educational Ministry	*Educational Statistics Yearbook of China. 1980-1986,1988, 1992* are downloaded from http://www.pinggu.org/bbs/
National, urban and rural population and birth rate for each year	• *China Statistics Yearbook 2008.* http://www.stats.gov.cn /tjsj/ndsj/2008/indexch.htm • *Statistics Summary for 55 years in China.* China Statistics Press	

Continue to the next page

Students by age and grade of primary and junior school: 2003-2007	• *Educational Statistics yearbook of China.2003-2007* edited by the Development and Plan Department of National Educational Ministry	
Forecasted national population and forecast national birth population: 2008-2020、2008-2020 forecast national birth population	• *The Report on China's National Strategy on Population Development（I）* by China Population Press	
The ratio of urbanization : 2008-2020	• *The Report on China's National Strategy on Population Development（II）* by China Population Press	

1.2 Micro-data

(1) Urban Household Survey (UHS)

The Urban Household Survey aims to study the conditions and standard of living of urban households. With the use of sampling techniques and daily accounting method, the survey collects data from non-agricultural households in different cities and counties. It records household information about income and consumption expenditure, demographic characteristics, work and employment, accommodation and other family related matters. This is a continuous, large scale social-economic survey, which covers from 1986 to 1997. 103 cities and 80 counties are included in the survey.

(2) China Health and Nutrition Survey (CHNS)

The China Health and Nutrition Survey was designed to examine the effects of the health, nutrition, and family planning policies and programs implemented by national and local governments and to see how the social and economic transformation of Chinese society is affecting the health and nutritional status of its population. The survey was conducted by an international team of researchers whose backgrounds include nutrition, public health, economics, sociology, Chinese studies, and demography. It is funded by National Institutes of Health (NIH). The CHNS is coordinated by Barry Popkin of the Carolina Population Center at the University of North Carolina. The CHNS is a collaborative project of the National Institute of Nutrition and Food Safety (INFS), the Chinese Center for Disease Control and Prevention (CCDC), and the University of North Carolina at Chapel Hill (UNC-CH). Dr. Fengying Zhai, Associate Director of the INFS, is the director of the Chinese group. Nine Provinces are covered by the survey: Guangxi, Guizhou, Heilongjiang, Henan, Hubei, Hunan, Jiangsu, Liaoning, Shandong. Four counties are selected in each province. In addition, the provincial capital and a lower income city were selected when feasible. The years surveyed include 1988, 1991, 1993, 1997, 2000, 2004 and 2006. CHNS 1989 included 3,795 households. 3,616 households, 3,441 households, 3,875 households, and 4,403 households participated in CHNS 1991, CHNS 1993, CHNS 1997 and CHNS 2000, respectively.

(3) Chinese Household Income Project (CHIP)

China's Income Distribution Project (1988-92), funded by Ford Foundation, was conducted by Institute of Economic Studies, Chinese Academy of Social Science in collaboration with some foreign scholars such as Keith Griffin, Carl Riskin and John Knight. China's income survey consists of two parts: urban and rural. The size of the urban sample is 9,009 households and the rural one is 10,258. Items surveyed include basic information of both the sample households and their members, focusing on income and wage, sources of income and household expenditure. For the rural households, information of their assets and debts, sale and consumption of their products, and purchase of production means were also collected. The 1995 survey records information on urban and rural household income and expenditure of that year. Because of the change in the economic structure in China, the questionnaire was redesigned to reflect this change. Provinces covered by the survey involve 28 provinces for the rural survey, excluding Xinjiang and Tibet and 10 provinces (Beijing, Shanxi, Liaoning, Jiangsu, Anhui, Henan, Hubei, Guangdong, Yunan and Gansu) for the urban residents. The years surveyed include 1988, 1995 and 2002.

2 Data processing

2.1 Basic population data

2.1.1 Census data

Due to direct registration and computer aggregation, the Census data do not take into account the left-out population.[20] The total populations from the 1982, 1990 and 2000 census data published at that time are slightly different from the population released in *China Statistics Yearbook 2008*. Thus, some adjustments need to be made to the population data by age, sex and educational attainment. The adjustment is described by the following method: the adjusted urban population by age, sex and educational attainment equals the urban population by age, sex and educational attainment from the census data times the ratio of total urban population released in *China Statistics Yearbook 2008* to the total urban population in the census data. A similar equation is applied to the rural population.

[20] See Zhang, Weimin and Hongyan Cui(2003),"The estimation accuracy of China Census 2000",*Population Research*, Vol.27, No.4 (July), pp.25-35.

2.1.2 1% sample data

We adjust the sample data to match the total rural and urban data. Urban population by age, sex and educational attainment is divided by urban sampling ratio, which is the ratio of urban sample population to urban total population released in *China Statistics Yearbook 2008.* The same method is applied to the rural data.

2.2 New enrollment

2.2.1 Educational category in China

There are six education levels in China: no schooling, primary school, junior middle school (including regular junior middle school and vocational junior middle school), senior middle school (including regular senior middle school, regular specialized middle school and vocational high school), college, and university and above. "College" and "university and above" were combined as "college and above" before 1990.

2.2.2 National enrollment data

The new enrollments by gender of primary school from 1985 to 1990 are not available, so it is assumed that the share of females in the new enrollments equals that in grade one. From 1980 to 1983, we have no information about the share of females in the new enrollments, so we use female share in new enrollment of the closest year. From 1980 to 2003, we only have new enrollment of college and university and the total females in college and university. To separate females in college and university, we assume that the proportion of female is the same as in college and university.

2.2.3 New enrollment data of urban and rural

The new enrollments by gender in urban and rural areas in each educational level are not available. We assume that the proportions of female in urban and rural equal the corresponding proportion at the national level. The new enrollments of specialized middle school are not separated by urban and rural. So we assume that the ratio of urban to rural new enrollments in specialized middle school is the same as that of regular senior middle school. From 2003 to 2007, the new enrollments of vocational high school are also not separated by urban and rural and the processing method is the same as above.

3 Imputation method

We use the perpetual inventory method to impute the population data.

3.1 Perpetual inventory method

The perpetual inventory formula is:

$$L(y,e,a,s) = L(y-1,e,a,s)\cdot(1-\delta(y,a,s)) + IF(y,e,a,s) - OF(y,e,a,s) + EX(y,e,a,s)$$

where $L(y,e,a,s)$ is the population at year y with education level e, age a and sex s. $\delta(y,a,s)$ is the mortality rate. $IF(y,e,a,s)$ is the inflow of population of age a, sex s and education level e in year y. $OF(y,e,a,s)$ represents the outflow of population of age a and sex s and education level e in year y. $EX(y,e,a,s)$ is a residual term.

$$IF(y,e,a,s) = \lambda(y,e,a,s)\cdot ERS(y,e,s)$$

$$OF(y,e,a,s) = \lambda(y,e+1,a,s)\cdot ERS(y,e+1,s)$$

ERS is the new enrollment of different education levels, λ is the age distribution of new enrollment of different education levels and

$$\sum_a \lambda(y,e,a,s)=1$$

3.2 Estimate the age distribution λ

3.2.1 Estimate the age distribution λ: using micro-data

The micro-data we use include CHNS (China Health and Nutrition Survey: 1989, 1991, 1993, 1997, 2000) and CHIP (Chinese Household Income Project: 1995). CHNS includes not only the age, gender of the individuals but also the grade if the individuals are in school, while CHIP only records the education level without grade. For this reason, we consider CHNS firstly when we estimate the age distribution of new enrollment.

3.2.1.1 Using CHNS data

a. The age distribution of the students at Grade 1 in primary school

Select the students at Grade 1 in primary school from the CHNS sample, and classify them according to age. The last two rows in Table A.1 show that the students at Grade 1 in primary school are mainly 5-10 years old, with the share over 95%. For simplicity and also for consistency with the age limits of other education levels, students aged less than 5 and over 10 are dropped from the sample. The age distribution is calculated for the students at Grade 1 in primary school aged 5-10 (Table A.2).

b. The age distribution of students at Grade 1 in junior middle school

The number of students at Grade 1 in junior middle school can be obtained by the same fashion, as shown in Table A.3. These students are mainly aged from 11 to 16, with the share over 95% except for 1993. In 1993, the number of students at Grade 1 in junior middle school is as large as 47, which is rare under the education framework of China, so they are dropped (Table A.4).

c. The age distribution of students at Grade 1 in senior middle school, college and university.

The number of the students at Grade 1 in senior middle school, college and university in CHNS sample is too small to estimate the age distributions. The number of students at Grade 1 in senior middle school is shown in Table A.5, and there are only 81 students at Grade 1 in college and university from 1989 to 2000 in CHNS sample.

3.2.1.2 Using CHIP95 data

Select the students in senior middle school (including professional schools), college and above (Table A.6). CHIP95 only records the education level, thus we do not know which grade the student is in. To estimate the age distribution for Grade 1, we assume the age distributions of students at each grade are the same as their Grade 1. Take the male students in senior middle school for example, as shown in Table A.7.

We also assume that the numbers of students at Grade 1, Grade 2 and Grade 3 are x, y, and z, respectively. We have

$$a\cdot x=26$$

$$b\cdot x+a\cdot y=72$$

$$c\cdot x+b\cdot y+a\cdot z=147$$

$$d\cdot x+c\cdot y+b\cdot z=203$$

$$e\cdot x+d\cdot y+c\cdot z=175$$

$$f\cdot x+e\cdot y+d\cdot z=61$$

$$f\cdot y+e\cdot z=60$$

$$f\cdot z=28$$

$$a+b+c+d+e+f=1$$

Solve these equations for the age distribution (*a,b,c,d,e,f,*). Similarly, we can derive the age distributions of female students at Grade 1 in senior middle school, male students at Grade 1 in college and university, and female students at Grade 1 in college and university. We present some results in Table A.8 and Table A.9.

3.2.2 Estimate the age distribution λ: using macro-data

We use the data in *China Educational Statistical Yearbook: 2003-2007* to estimate the age distribution of new enrollments.

We have the data of new enrollment of primary school by age and the data of new enrollment of junior middle school by age and grade from 2003 to 2007.

For primary school, we assume that males and females have the same age distribution.

For junior middle school, we assume that the age distribution of Grade 1 students is the same as that of new enrollment. Then we assume that males and females have the same age distribution.

For senior middle school, we assume Grade 3 students in junior middle school have the same age distribution as that of new entrants to senior middle school in the same year. Then we assume that males and females have the same age distribution. For example, in 2004 the age distribution of new entrants to senior middle school is the same as that of Grade 3 students in junior middle school (TableA.10).

For university, we assume that the age distribution of new entrants to university is the same as that of Grade 1 students in senior middle school three years ago. For example, in 2007, the age distribution of new entrants to university is the same as that of Grade 1 students of senior middle school in 2004. See Table A.11.

Using the method above, we can get the age distribution of enrollment of each educational level (Table A.12). Here males and females have the same age distribution.

3.3 Method of imputing population data: 1985-2005

When adopting the perpetual inventory method to estimate the urban and rural population, we ignore migrants between urban and rural China. To take these migrants into account, we make the following adjustments. For example, from 1982 to 1990, we get the estimated 1990 population data by gender, education and age from the perpetual inventory method. The actual 1990 population by gender, education and age subtracted the estimated 1990 population by gender, education and age gives the net migrants between urban and rural China in these eight years. We assume that the number of immigrants in each year is the same, and then we add the average difference to the estimated population data.

3.4 Method of imputing population data: 2006-2020

With the population by age, gender and education level of 2005 as the benchmark, we use the perpetual inventory method to obtain preliminary estimates, and then adjust the sum of population estimated to match data released in *China Statistics Yearbook 2008*, and then forecast population from 2008 to 2020.

The method of adjustment is, we use the total population reported in *China Statistics Yearbook 2008*, minus the sum of the estimated population retrieve the difference. Then we add the difference back to the estimated population data according to the 2005 structure of the population by age, gender and education level.

When it comes to estimating the enrollment data, we assume that the enrollment rate of the population of a certain sex, age and education level from 2008 to 2020 equals that of the 2005 population. For example, the rate of male population of 15 years old of junior middle school in 2004 divided by male entrants of 16 years old of senior middle school in 2005 is defined as the enrollment rate. Thus we get the enrollment rate by age, gender and education level. When we calculate the number of population in college and university of rural areas, we assume that the change of each year equals that from 2004 to 2005.

4 Some specific problems

4.1 National, rural and urban population at age zero: 1985-2007

4.1.1 National population at age zero

The total populations at the year end and the birth rates for each year are obtained from Table 3-1 'Population and Its Composition' and Table 3-2 'Birth Rate, Death Rate and Natural Growth Rate of Population' in *China Statistic Yearbook 2008*. We assume that the population at the beginning of a given year equals that at the end of the previous year. Thus, the average of the populations at the end of the given year and the previous year is the average population of the given year. The product of the average population and the corresponding birth rate gives the new-born population. Multiplying the new-born population and the survival rate of those aged zero at the corresponding year gives the population at age zero at the end of the year.

(Definition[21]: birth rate, also called gross birth rate, refers to the ratio of the new-born population in a given region during a given period, usually one year, and the average population of the same period. The birth rate here is yearly birth rate, which is calculated from the following equation: Birth rate = (new-born population/average population)* 1000‰,

[21] From *China Statistics Yearbook 2008*.

where new-born population is the number of the new-born babies who are alive when they are detached from the mothers no matter how long they have been in their mother's body. Average population is the average of the populations at the beginning and at the end of the year, or the population at the middle of the year.)

4.1.2 Rural and urban population at age zero

The data used include: total national population for each year from 1983 to 2007, birth rate for each year from 1983 to 2007, national, rural and urban population by age and gender from the population sampling surveys for 1987 and each year from 1989 to 2007.

The share of urban population at age zero in the national population at age zero can be calculated from these sampling data, and this share is assumed to be the true share, i.e. multiplying it with the national population at age zero produces the urban population at age zero. Further, the gender ratio from the sampling data is also assumed to be true, thus we can divide the urban population at age zero into the two genders. Similar steps are used for the rural population at age zero.

Since there is no population sampling data for 1983-1986 and 1988, we assume the numbers of those aged 1, 3, 4, 5, 6 equals the new-born population for 1988, 1986, 1985, 1984 and 1983 respectively with the sampling weights adjusted. Migration between urban and rural regions is neglected here.

4.2 The application of forecasted population data: 2008-2020

4.2.1 National population and birth population

The following assumptions are made when forecasting the national population:[22]

[22]See Jiang Zhenghua(2007),The Report on China's National Strategy on Population Development（I）,China Population Press, pp.1001-1128

a. From 2004 the non-agricultural population's PTFR=1.18, agricultural population's PTFR=1.88

b. Only single-child couples can give birth to a second child, and the second child's PTFR（2）=0.95. Assume that single-and non-single child marry randomly, the third child's PTFR follows the original plan.

c. It is assumed that people transfer from agricultural population to non-agricultural population during 2000 to 2005. At the end of 2000, the ratio of the non-agricultural population to the total population is 24.7%, and it reaches 55.0% at the end of 2050.

See Table A.13.

4.2.2 The rate of urbanization

It is assumed that the rate of urbanization increase by 1% per year from 2007 to 2020. The rate of urbanization is the ratio of urban population to the total population. See Table A.14.

4.3 Urban and rural population aged zero from 2008 to 2020

We have data of forecast national birth population from 2008-2020, but we need to separate it by urban and rural. We assume that the ratio of urban birth population to rural birth population from 2008-2020 is equal to the ratio of 2007. We also assume that the ratio of the number of the male new-born to the number of the female new-born is equal to that of 2007. Then we convert the birth population to the population aged zero by using the death rate of those aged zero which also equals that of 2007.

4.4 The death rate of those aged 65 and over

4.4.1 The death rate of those aged 65 and over: 1985-2007

When imputing the population by age, gender and education level with perpetual inventory method, the number of those aged 65 and over should be multiplied by (1-death rate), where the death rate is calculated in the following way. With the population and the death rate, both by age and gender, from the population sampling data for each year, the number of deaths of those aged 65 and over for each year can be calculated, and dividing it by the corresponding total population gives the death rate of those aged 65 and over. Since there is no population sampling data for 1983-1986 and 1988, the death rate of the closest year is used.

4.4.2 The death rate of those aged 65 and over: 2008-2020

The death rate of those aged 65 and over from 2008 to 2020 equals that of 2007.

4.5 Application of the age distributions of every education level for each year

The age distributions are obtained from the macro- and micro-level data, and the enrollment numbers for each year are used with adjustments. They change over time, but do not vary between urban and rural regions.

4.6 STATA programming

The imputation process is realized by a STATA program, which includes negative numbers adjustments.

Tables and Figures of Appendix A

Table A.1 Number of students at Grade 1 in primary school in CHNS sample

Age	1989		1991		1993		1997		2000	
	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female
4	1									
5	7	5	13	8	3	3	11	6	5	3
6	48	39	32	30	14	13	31	37	12	9
7	67	64	41	40	21	9	50	47	22	12
8	47	23	24	12	5	4	23	7	6	3
9	6	4	10	6	3	2	3	1		4
10	3	2	2	3	2	3	1	1		1
11							1		2	
12	1	1	2	1	1					
13	1						1			
14	1		1			1		1		1
15				1						
16							1			
25							1			
Total	182	138	125	101	49	35	123	100	47	33
Those aged 5-10	178	137	122	99	48	34	119	99	45	32
The share of those aged 5-10	0.978	0.993	0.976	0.98	0.98	0.971	0.967	0.99	0.957	0.97

Table A.2 Age distribution of students at Grade 1 in primary school in CHNS sample

Age	1989		1991		1993		1997		2000	
	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female
5	0.0393	0.0365	0.1066	0.0808	0.0625	0.0882	0.0924	0.0606	0.1111	0.0938
6	0.2697	0.2847	0.2623	0.3030	0.2917	0.3824	0.2605	0.3737	0.2667	0.2813
7	0.3764	0.4672	0.3361	0.4040	0.4375	0.2647	0.4202	0.4747	0.4889	0.3750
8	0.2640	0.1679	0.1967	0.1212	0.1042	0.1176	0.1933	0.0707	0.1333	0.0938
9	0.0337	0.0292	0.0820	0.0606	0.0625	0.0588	0.0252	0.0101	0.0000	0.1250
10	0.0169	0.0146	0.0164	0.0303	0.0417	0.0882	0.0084	0.0101	0.0000	0.0313
Total	1	1	1	1	1	1	1	1	1	1

Table A.3 Number of students at Grade 1 in junior middle school in CHNS sample

Age	1989		1991		1993		1997		2000	
	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female
6					1					
7					8	7				
8					4	12				
9	1				9	6				
10					2		1		2	
11	5	1	5	8	7	8	6	11	16	10
12	16	21	24	23	28	31	26	19	51	38
13	36	32	22	30	34	30	41	43	56	40
14	35	21	22	28	25	22	20	19	23	12
15	18	8	16	11	11	6	7	4	3	3
16	8	4	10	1	1	1	1	2	1	1
17	1		4		1	3	1		1	
18				1		1		1		
19		1								1
21				1						
22							1			
35	1									
36		1								
38				1						
45	1					1				
63								1		
Total	122	89	103	104	131	128	104	100	153	105
Those aged 11-16	118	87	99	101	106	98	101	98	150	104
The share of those aged 11-16	0.97	0.98	0.96	0.97	0.81	0.77	0.97	0.98	0.98	0.99

Table A.4 Age distribution of students at Grade 1 in junior middle school in CHNS sample

Age	1989		1991		1993		1997		2000	
	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female
11	0.0424	0.0115	0.0505	0.0792	0.0660	0.0816	0.0594	0.1122	0.1067	0.0962
12	0.1356	0.2414	0.2424	0.2277	0.2642	0.3163	0.2574	0.1939	0.3400	0.3654
13	0.3051	0.3678	0.2222	0.2970	0.3208	0.3061	0.4059	0.4388	0.3733	0.3846
14	0.2966	0.2414	0.2222	0.2772	0.2358	0.2245	0.1980	0.1939	0.1533	0.1154
15	0.1525	0.0920	0.1616	0.1089	0.1038	0.0612	0.0693	0.0408	0.0200	0.0288
16	0.0678	0.0460	0.1010	0.0099	0.0094	0.0102	0.0099	0.0204	0.0067	0.0096
Total	1	1	1	1	1	1	1	1	1	1

Table A.5 Number of students at Grade 1 in senior middle school in CHNS sample, with professional school included

Age	1989		1991		1993		1997		2000	
	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female
11					1	1				
12						2				
13			1		1					
14	1	2	2		1		1	5	1	4
15	6	8	9	6	10	11	13	13	7	9
16	10	5	9	7	6	10	19	14	16	20
17	5	5	5	5	6	10	4	10	15	9
18	1	1	1		4	1	1	3	3	5
19	1	1	2			2	1	2		3
20									1	1
21			1							
28								1		
Total	24	22	30	18	29	37	39	48	43	51

Table A.6 Number of Students in senior middle school and above in CHIP95 sample

Age	Senior middle school (including professional schools)	Senior middle school (including professional schools)	College and higher	College and higher
	Male	Female	Male	Female
1	1			
2		1		
3		1		
4	1			
5				1
6	2	1	1	1
7			2	3
8		3	5	3
9	1	1	1	
10	6	2	1	1
11	2	3		1
12	5	4	4	
13	14	16		3
14	26	23	1	1
15	72	78	1	4
16	147	176	2	4
17	203	162	6	10

Continue to the next page

18	175	164	17	20
19	61	86	26	22
20	60	45	34	26
21	28	23	21	19
22	13	11	16	9
23	6	3	11	4
24	2	2	3	5
25		2	5	
26				1
27			1	
28	1			
31		1		
38		1		
40		1		
88		1		
Total	826	811	158	138
Of which:	age 14-21		age 17-24	
Number of students	772	757	134	115
share	0.9346	0.9334	0.8481	0.8333

Table A.7 The assumption that the age distributions of students at each grade are the same as their Grade 1

Age	Grade 1	Grade 2	Grade 3
14	a		
15	b	a	
16	c	b	a
17	d	c	b
18	e	d	c
19	f	e	d
20		f	e
21			f

Table A.8 Age distribution of male students at Grade 1 in senior middle school

Age	Share
15	0.273
16	0.351
17	0.158
18	0.144
19	0.085

Table A.9 Age distributions for Grade 1 in college and university

Age	Male	Female
17	0.044	0.044
18	0.423	0.423
19	0.438	0.438
20	0.082	0.082
21	0.013	0.013

TableA.10 Age distribution of new entrants of senior middle school of 2004

Age	Grade three students in junior middle school	Proportion
11 and below	21	0.000001
12	2185	0.000098
13	79869	0.003586
14	1279586	0.057452
15	8893796	0.399322
16	9785227	0.439346
17	1899324	0.085278
18	293469	0.013176
19 and above	38789	0.001742
Total	22272266	1.000000

Table A.11 Age distribution of new entrants of university of 2007

Age	Proportion
14 and below	0.000001
15	0.000098
16	0.003586
17	0.057452
18	0.399322
19	0.439346
20	0.085278
21	0.013176
22 and above	0.001742
Total	1.000000

Table A.12 Age distribution of enrollment of each educational level of 2007

Age	2007				
	Illiterate to primary school	Primary school to junior middle school	Junior middle school to senior middle school	Senior middle school to college	Senior middle school to university
	Proportion	Proportion	Proportion	Proportion	Proportion
5	0.029				
6	0.624				
7	0.325				
8	0.018				
9	0.003				
10	0.001	0.001			
11		0.041			
12		0.445			
13		0.415	0.002		
14		0.079	0.006		
15		0.016	0.447		
16		0.003	0.440	0.004	0.004
17		0.001	0.087	0.058	0.058
18			0.015	0.399	0.399
19			0.003	0.439	0.439
20				0.085	0.085
21				0.013	0.013
22				0.002	0.002
Sum	1	1	1	1	1

Table A.13 Forecast national population and birth population, in millions

Year	National population	New-born population
2008	1,328.70	16.71
2009	1,336.00	17.08
2010	1,343.50	17.49
2011	1,351.40	17.89
2012	1,359.30	18.17
2013	1,367.30	18.28
2014	1,375.10	18.23
2015	1,382.50	18.02
2016	1,389.50	17.69
2017	1,396.00	17.29
2018	1,401.90	16.86
2019	1,407.20	16.41
2020	1,411.90	15.97

Table A.14 Rate of urbanization from 2008-2020

Year	Rate of urbanization
2008	0.4594
2009	0.4694
2010	0.4794
2011	0.4894
2012	0.4994
2013	0.5094
2014	0.5194
2015	0.5294
2016	0.5394
2017	0.5494
2018	0.5594
2019	0.5694
2020	0.5794

Appendix B Mincer parameters

Main Formula:

$$\ln(inc) = \alpha + \beta \cdot e + \gamma \cdot exp + \delta \cdot exp^2 + u \tag{1}$$

1 Outline of Samples and Methods

1.1 Surveys

(1) The annual Urban Household Survey;

(2) Chinese Health and Nutrition Survey.

1.2 Components of Income

(1) Regular and non-regular income;

(2) Other cash income from work unit;

(3) Estimated market value of items received;

(4) All kinds of subsidies;

(5) Part of income of rural individuals is derived from household income according to working-hour share.

1.3 Work experience

exp = age–e– 6.

1.4 Sample selection criteria

(1) 16-60 years old for males, and 16-55 years old for females;

(2) Must have information on income and educational attainment;

(3) Exclude: students, retirees, waiting for job, disabled, waiting to enter a school, homemakers.

1.5 Imputation method

(1) To make all parameters comparable, we use UHS to obtain all urban parameters and CHNS to obtain urban/rural ratios of parameters, and to derive rural parameter estimates as: rural parameters = urban parameters/ ratio.

For example, from UHS1989, we can get α^u89(UHS); from CHNS1989, we can get α^u89(CHNS) and α^r89(CHNS); then, the ratio for intercept(CHNS) = α^u89(CHNS)/α^r89(CHNS); finally, we estimate 1989 parameter for rural group, α^r89 = α^u89(UHS) /ratio(CHNS). We get β^r89, γ^r89, δ^r89 in a similar manner. Note: α^r89 indicates the parameter of year 1989 for rural population.

(2) For urban parameters, we can use UHS to get urban parameters for the years 1986-1997, but we actually need parameters for the years 1985-2007. We fit a time trend model for each of the parameters and obtain parameter estimates for the missing years as the fitted values of the time trend model.

(3) We can obtain urban/rural parameter ratios only for 1989, 1991, 1993, 1997, and 2000, which are then used to fit a time trend model for each parameter. We obtain the ratios for the rest of the years as the fitted values of the time trend models.

1.6 Parameter α

$$\ln(y) = a0 + a1 * e + a2 * exp + a3 * exp^2$$

$\hat{y}$ is not equal to $e^{\widehat{\ln(y)}}$, rather $\hat{y} = \alpha * e^{\widehat{\ln(y)}}$, where α is an adjustment factor. We estimate it as follows:

(1) Obtain $\widehat{\ln(y_i)}$ from the regression of ln (yi) on all right-hand-side variables.

(2) Obtain $\hat{m}_i = e^{\widehat{\ln(y_i)}}$.

(3) Regress yi on $\hat{m}_i$ without the intercept: $\hat{y} = \alpha * \hat{m}_i$ and keep α.

(4) For given values e= c1, exp = c2, exp^2 = c3, obtain $\widehat{\ln(y)}$.

(5) $\hat{y} = \alpha * \widehat{\ln(y)}$.

For the urban population, for the years we have data, we can get α_value directly. For the years we do not have survey data, we use the α_value for the most recent year for which α is estimated using the procedure described above.

For rural, we use the same α_value as urban in the same year.

1.7 Forecasting values

While there is evidence suggesting these parameters follow a time trend up to recent years, it is hard to determine if similar trend will continue beyond 2008. Therefore, we assume the fitted urban/rural parameter ratios as well as imputed parameter values for the period 2008 to 2020 to be constant and equal to the values of their counterparts in 2007.

2 Data availability

The data used for estimating the parameters of the earnings equation come from two well-known household surveys in China. The first is the annual Urban Household Survey (UHS) conducted by the National Statistical Bureau of China over the period 1986 to 1997. It records household information about income and consumption expenditure, demographic characteristics, work and employment, accommodation and other family related matters. This is a continuous, large scale social-economic survey, which covers from 1986 to 1997. 103 cities and 80 counties are included in the survey. We use this data set to estimate the parameters of equation (1) for each gender of the urban population by year and fit these estimates in a linear or exponential time trend model. We then use the fitted time trend model to generate the imputed parameters of the earnings equation for the urban population for the period 1985 through 2020.

The second data set we use is the China Health and Nutrition Survey (CHNS). Nine Provinces are covered by the survey: Guangxi, Guizhou, Heilongjiang, Henan, Hubei, Hunan, Jiangsu, Liaoning, Shandong, four counties are selected in each province. In addition, the provincial capital and a lower income city were selected when feasible. The years surveyed include 1988, 1991, 1993, 1997, 2000, 2004 and 2006. CHNS 1989 included 3,795 households. 3,616 households, 3,441 households, 3,875 households, and 4,403 households participated in CHNS 1991, CHNS1993, CHNS 1997 and CHNS 2000, respectively. Here we don't use the data for 2004 and 2006, because the questionnaires for these two years changed a lot form before, then it might make the results not comparable. This survey covers both the urban and rural population. We use CHNS to obtain earnings equation parameter estimates by year for each gender and separately for the rural and urban population. We calculate the urban-to-rural ratio for each of these parameters. We then use the ratio to fit a time trend model, which is used to generate fitted values of the ratio over the period 1985 to 2020. We use the fitted values along with the imputed parameters for the urban population to derive the imputed parameters for the rural population over the period 1985 to 2020.

We didn't use the CHIP (Chinese Household Income Project) data. In this data set, the size of the urban sample is 9,009 households and the rural one is 10,258. Items surveyed include basic information of both the sample households and their members, focusing on income and wage, sources of income and household expenditure. For the rural households, information of their assets and debts, sale and consumption of their products, and purchase of production means were also collected. Provinces covered by the survey involve 28 provinces for the rural survey, excluding Xinjiang and Tibet and 10 provinces (Beijing, Shanxi, Liaoning, Jiangsu, Anhui, Henan, Hubei, Guangdong, Yunan and Gansu) for the urban residents. The years surveyed include 1988, 1995 and 2002.There are two reasons that we don’t use this data set. From one aspect, this data set only covers three years of 1988, 1995 and 2002, which are 7-year apart from each other; in addition, the survey questions and the people who are surveyed changed over the three sample years. These might make the results difficult to compare. As a result, we finally choose UHS and CHNS to estimate the final parameters for 1985-2020. The distribution of the two data sets across years is shown in Table B.1.

3 Description of each data set

3.1 CHNS

3.1.1 Income variables

Income includes three parts: wage, subsidies and agricultural income.

3.1.1.1 Wage

Wage consists of two components: wage payments according to finished piece of work and wage payments according to hours worked.

Variables：

(1)1989

wage = c5 * c3 * 52 or wage = c6 * c7 * 52

c3: average days/week worked

c5: wage for a usual day’s work

c6: wage per piece of completed work

c7: average number of pieces completed/week

(2) 1991-2006

wage = c3 * c8

c3: months worked last year

c8: average monthly wage

3.1.1.2 Subsidies and other kinds of income

Subsidies include food subsidies, one-child subsidy, health subsidy, bathing & haircutting subsidy, books & newspaper subsidy, etc. Other kinds of income include bonuses, gifts, coupons, and in-kind income.

Variables：

totalsub = (I9 + I10 + I11 + I12 + I13 + I14) * 12 + I19 + I21

I9: monthly food subsidy

I10: monthly one-child subsidy

I11: monthly health subsidy

I12: monthly bath/haircut subsidy

I13: monthly book/newspaper subsidy

I14: monthly other subsidies

I19: 12-mo total salary bonuses

I21: value of gift/coupons (only for 1989)

3.1.1.3 Agricultural income

Agricultural income includes incomes from five sources: gardening, farming, livestock raising, fishing, and small handicraft & commercial household businesses. These incomes come from either collective or household businesses or both.

(1)Gardening

Household

gardhhdinc = (D5 - D7)+ D6 * 12

D5: sale of home-plot produce

D6: saving from consuming home-plot produce

D7: money spent on home garden last year

(2)Farming

Household

farmhhdinc = E15 + E17 + E19 - E13

E13: 12-mo. amount spent on raising this crop

E15: government price/kilo for crop last year

E17: 12-mo. amount received if kept-farm-crop sold

E19: 12-mo. amount received if given-farm-crop sold

Collective

farmcltinc = E7 + E9

E7: 12-mo. amount from collective farm work

E9: amount of durable goods worth

(3)Livestock rising

Household

livestockhhdinc = F17 + F19 + F21 + F15- F14

F14: 12-mo. amount spent on animal operation

F15: 12-mo. amount saved using homegrown feed

F17: amount received from sale of animal products

F19: value if kept animal products sold

F21: value if animal products given away sold

Collective

livestockcltinc = F7 + F9;

F7: amount received for animal work in collective

F9: market value of animals received, if sold

(4)Fishing

Household

fishhhdinc = G11 + G13 + G15 - G16

G11: revenue from fish sales (In 1989 and 1991, G11 is measured by month, other years it is measured by year)

G13: amount earned if HH fish sold

G15: amount earned if gift fish sold

G16: operating expenses of fish business

Collective

fishcltinc = G7 + G9

G7: amount received from collective fishing

G9: amount received if fish from collective sold

(5)Small handicraft & commercial household business

Household

commercialinc = 12 * (H3 - H4);

H3: average monthly revenues, home business

H4: average monthly expenses, home business

(In1989, H3/H4 are measured by week，other years they are measured by month)

3.1.2 Distributing income

3.1.2.1 Reason

Agricultural income includes incomes from five sources: gardening, farming, livestock raising, fishing, and small handicraft & commercial household businesses. These incomes come from either collective or household businesses or both.

3.1.2.2 Rules

We assume each individual's contribution to the household income is proportional to his or her share of time allocated to five activities: gardening, farming, raising livestock, fishing and small handicraft & commercial household business. First, we add up all working hours of all family members in each of these activities. Second, we calculate the working hour share of each member in the family's total hours. Third, we multiply the household income by the share to approximate individual income. Finally, we add up individual income from the four categories for each family member.

3.1.3 Defining years of schooling

Level	*e*
None	0
Completed primary school	6
Junior middle school degree	9
Senior middle school degree	12
Middle technical, professional , or vocational degree	11
3- or 4- year college degree	16
Master's degree or above	18

3.1.4 Sample selection criteria

(1) Males of 16 to 60 years of age and females of 16 to 55 years of age;

(2) Exclude individuals who failed to provide information on wage and educational attainment, those who are self-employed or business owners;

(3) Incomes from secondary work are not included.

3.2 UHS

3.2.1 Definition of income

(1)1986-1987

Monthly wage: u010--u080

Yearly income=monthly wage*12

Hourly wage=yearly income/(52*5*8)

(2)1988-1991

Regular wages: v0012-v7012

Other income from work unit: v0019-v7019

Income of employees of individual enterprise: v0022-v7022

Yearly income = normal wage + other income from work unit + income of employees of individual enterprise

Hourly wage=yearly income / (52*5*8)

(3)1992

Regular wages: vp113-vp813

Other income from work unit: vp120-vp820

Income of employees of individual enterprise: vp122-vp822

Yearly income = normal wage + other income from work unit + income of employees of individual enterprise

Hourly wage=yearly income / (52*5*8)

(4)1993-1997

Wages: x13/x50/x87/ x124/ x161/ x198/ x235/ x272

Operating profit: x22/x59/x96/ x133/ x170/ x207/ x244/ x281

Yearly income = wages+ operating profit

3.2.2 Years of schooling

(1) 1986-1991

Level	*e*
College	16
Professional school	11
Senior middle school	12
Junior middle school	9
Primary school	6
Others	0

(2) 1992-1997

Level	*e*
College	16
Community college	15
Professional school	11
Senior middle school	12
Junior middle school	9
Primary school	6
Others	0

3.2.3 Principles of selecting samples

(1) Include male individuals of 16 to 60 years old and female individuals from 16 to 55 years old;

(2) Omit individuals whose value of regular wage is missing, individuals who failed to report education information;

(3) Omit individuals who are self-employed, the short term contract workers, retired, job seekers, disabled, homemakers, students in school, workers waiting for job assignment, students waiting to enter a higher school, etc.

4 Imputing parameters

4.1 Imputation method of urban parameters

4.1.1 Parameter estimates based on UHS 1986 to 1997

We use the UHS data to estimate the earnings equation for each gender by year. Table B.2 contains means and standard deviations of each variable. The estimates are reported in Table B.3.

4.1.2 General idea about the method used

We use the parameter estimates over the period 1988-1997 to fit a time trend model, and then get the fitted values of each parameter by gender for the years 1985-2007. These fitted values are the final urban imputed parameters.

4.1.3 Functional fitting

4.1.3.1 Function form

We treat $\alpha, \beta, \gamma, \delta$ separately and use the parameters of each group as the dependent variable and use time (i.e., year) as the independent variable.

For α and β, we use the linear time trend model. We rely on R^2 and AIC values and SIC values in choosing among alternative regression models. The regression equation is: $Y = a_0 + a_1 * time + u$.

For γ and δ, we use the exponential time trend model. We rely on R^2 and AIC values and SIC values in choosing among alternative regression models. The regression equation is: $\ln(Y) = a_0 + a_1 * time + u$.

4.1.3.2 Assumptions

For α and β, we assume that they increase or decrease at a constant rate each year. Taking the α_male as an example, we assume that the intercept increases at the growth rate of a1 per year.

For γ and δ, we assume that they increase or decrease at a constant rate in percentage terms per year. Taking γ_male as an example, we assume that the coefficient of exp increases a1*100% per year.

4.1.3.3 Some special treatments

(1) As the coefficient of exp^2 is negative, we use its absolute value as dependent variable and use the negative of the fitted value as its imputed value.

(2) From Figures B.1-8, we can see that the estimates for 1986 and 1987 deviate substantially from the time trend. We exclude them when fitting time trend models.

4.1.3.4 Data and figures

Figures B.9-16 show the parameter estimates of each group and the sample regression lines of the time trend models. The fitted values of the time trend models are the values of our imputed parameters for the period 1985 to 2007.

4.1.3.5 Imputed urban parameter values by gender and year

Assuming that the imputed parameters remain constant after 2008 and equal to their values in 2007, we obtain the imputed parameters as shown in Table B.4.

4.2 Imputation method of ratio parameters

4.2.1 Urban-to-rural ratios of parameter estimates

We use CHNS data to get the urban and rural estimates of the earnings equation and then obtain the urban-to-rural ratios of parameter estimates. Table B.5 contains means and standard deviations of each variable.The specific urban and rural estimates and the corresponding ratios are showed in Table B.6.

4.2.2 Fitting methods

4.2.2.1 Function form

We treat $\alpha, \beta, \gamma, \delta$ separately and use the parameters of each group as the dependent variable and time as the independent variable. We fit an exponential model for each of them. The regression model is: ln(Y)= $a_0 + a_1$ * time + u.

4.2.2.2 Assumptions

We assume that the urban-to-rural ratios for $\alpha, \beta, \gamma, \delta$ increase or decrease at a constant rate in percentage terms per year. Taking the ratio of α_male as an example, we assume that the ratio increases by a_1*100% per year.

4.2.2.3 Data and figures

Figures B.17-24 show the sample regressions lines of the exponential trend models. The fitted values are the imputed ratios for the period 1985 to 2007.

4.2.3 Final ratio parameters

We assume the ratios remain constant for the period 2008 to 2020 and equal to the values of their counterparts in 2007. Table B.7 shows the imputed ratios for the period 1985 to 2020.

4.3 Imputed parameters for the rural population

Rural parameters = urban parameters/ratio, Table B.8 shows the imputed values of the rural parameters.

4.4 α_value

Table B.9 shows α_value for each year.

Tables and Figures of Appendix B

Table B.1 Data availability across years

Year	CHNS	UHS
1985		
1986		U
1987		U
1988		U
1989	U/R	U
1990		U
1991	U/R	U
1992		U
1993	U/R	U
1994		U
1995		U
1996		U
1997	U/R	U
1998		
1999		
2000	U/R	
2001		
2002		
2003		
2004	U/R	
2005		
2006	U/R	
2007		

Note: CHNS: China Health and Nutrition Survey
UHS: Urban Household Survey

Table B.2 Summary statistics: UHS Samples

year	variables	male		female	
		mean	s.d.	mean	s.d.
1986	inc.	1297.96	483.99	1024.33	408.31
	e	9.90	2.71	9.32	2.44
	exp	20.93	9.74	17.50	7.48
	exp^2	532.75	451.62	361.96	278.57

Continue to the next page

1987	inc.	1371.29	554.52	1095.88	2.60
	e	10.11	2.75	9.44	8.05
	exp	21.99	10.17	18.08	304.86
	exp^2	586.97	476.19	391.73	499.15
1988	inc.	1305.24	572.86	1084.10	485.04
	e	10.81	2.92	9.99	2.72
	exp	20.46	10.79	17.78	9.27
	exp^2	534.94	462.39	401.93	339.21
1989	inc.	1271.55	588.98	1061.46	508.92
	e	10.96	2.95	10.15	2.67
	exp	20.68	10.85	18.15	9.25
	exp^2	545.26	463.82	414.79	340.43
1990	inc.	1391.31	616.28	1168.13	537.63
	e	11.12	2.91	10.33	2.68
	exp	21.08	10.73	18.35	9.20
	exp^2	559.29	465.08	421.40	339.07
1991	inc.	1459.93	642.87	1243.88	560.47
	e	11.28	2.93	10.54	2.63
	exp	20.57	10.44	18.09	8.92
	exp^2	532.10	450.88	406.64	329.27
1992	inc.	1665.07	847.26	1408.29	684.67
	e	11.43	2.75	10.75	2.53
	exp	20.89	10.47	18.47	8.91
	exp^2	545.81	454.58	420.52	331.40
1993	inc.	1723.47	1101.08	1457.79	886.08
	e	11.41	2.70	10.79	2.52
	exp	21.19	10.47	18.83	8.94
	exp^2	558.60	455.37	434.34	332.45
1994	inc.	1936.37	1298.04	1600.68	1079.34
	e	11.54	2.75	10.96	2.46
	exp	21.01	10.42	18.66	8.95
	exp^2	549.83	453.84	428.30	335.27
1995	inc.	2028.32	1278.67	1697.88	1095.80
	e	11.62	2.71	11.00	2.47
	exp	21.27	10.17	18.92	8.81
	exp^2	555.58	442.11	435.46	330.75

Continue to the next page

1996	inc.	2049.76	1434.03	1718.10	1273.71
	e	11.65	2.68	11.11	2.40
	exp	21.60	10.22	19.26	8.86
	exp^2	571.04	446.94	449.47	334.93
1997	inc.	2307.20	1692.37	1912.28	1488.65
	e	11.67	2.67	11.14	2.40
	exp	21.80	10.05	19.47	8.90
	exp^2	576.19	439.15	458.28	338.97

Table B.3 Estimates of the earnings equation: UHS Samples

year	male				female			
	α	β	γ	δ	α	β	γ	δ
1986	6.23576	0.01733	0.04990	-0.00068	5.93734	0.04191	0.04009	-0.00052
1987	6.30749	0.01972	0.04401	-0.00057	6.21450	0.03498	0.02692	-0.00035
1988	5.82832	0.03011	0.07377	-0.00114	5.45430	0.05372	0.08222	-0.00149
1989	5.77330	0.03439	0.07000	-0.00104	5.48916	0.05408	0.07291	-0.00126
1990	5.90239	0.03487	0.06492	-0.00094	5.59197	0.05612	0.06922	-0.00118
1991	6.04919	0.03434	0.05797	-0.00083	5.73593	0.05424	0.06254	-0.00104
1992	6.11499	0.04256	0.05333	-0.00074	5.71777	0.06484	0.06199	-0.00104
1993	6.04489	0.04847	0.05139	-0.00069	5.67653	0.07269	0.05478	-0.00085
1994	5.96259	0.06311	0.04913	-0.00062	5.47777	0.09354	0.05503	-0.00085
1995	6.08869	0.06006	0.04471	-0.00053	5.61289	0.08757	0.05414	-0.00082
1996	5.94992	0.06845	0.04642	-0.00055	5.62366	0.09123	0.04320	-0.00054
1997	6.01672	0.07218	0.04450	-0.00052	5.51068	0.10781	0.04197	-0.00051

Table B.4 Imputed earnings equation parameters for the urban population, 1985 to 2020

year	male				female			
	α	β	γ	δ	α	β	γ	δ
1985	5.81248	0.01089	0.08555	-0.00147	5.55553	0.02677	0.09859	-0.00209
1986	5.83390	0.01595	0.08061	-0.00134	5.56000	0.03301	0.09198	-0.00187
1987	5.85532	0.02101	0.07595	-0.00122	5.56447	0.03926	0.08581	-0.00167
1988	5.87673	0.02608	0.07156	-0.00111	5.56894	0.04550	0.08006	-0.00150
1989	5.89815	0.03114	0.06742	-0.00102	5.57342	0.05174	0.07469	-0.00134
1990	5.91956	0.03620	0.06353	-0.00093	5.57789	0.05798	0.06968	-0.00120
1991	5.94098	0.04126	0.05986	-0.00084	5.58236	0.06422	0.06501	-0.00107
1992	5.96239	0.04632	0.05640	-0.00077	5.58683	0.07046	0.06065	-0.00096
1993	5.98381	0.05138	0.05314	-0.00070	5.59130	0.07670	0.05658	-0.00086

Continue to the next page

1994	6.00522	0.05645	0.05007	-0.00064	5.59577	0.08295	0.05279	-0.00077
1995	6.02664	0.06151	0.04717	-0.00058	5.60024	0.08919	0.04925	-0.00069
1996	6.04805	0.06657	0.04445	-0.00053	5.60472	0.09543	0.04595	-0.00062
1997	6.06947	0.07163	0.04188	-0.00048	5.60919	0.10167	0.04287	-0.00055
1998	6.09088	0.07669	0.03946	-0.00044	5.61366	0.10791	0.03999	-0.00049
1999	6.11230	0.08176	0.03718	-0.00040	5.61813	0.11415	0.03731	-0.00044
2000	6.13372	0.08682	0.03503	-0.00037	5.62260	0.12040	0.03481	-0.00040
2001	6.15513	0.09188	0.03300	-0.00033	5.62707	0.12664	0.03248	-0.00035
2002	6.17655	0.09694	0.03110	-0.00030	5.63155	0.13288	0.03030	-0.00032
2003	6.19796	0.10200	0.02930	-0.00028	5.63602	0.13912	0.02827	-0.00028
2004	6.21938	0.10707	0.02761	-0.00025	5.64049	0.14536	0.02637	-0.00025
2005	6.24079	0.11213	0.02601	-0.00023	5.64496	0.15160	0.02460	-0.00023
2006	6.26221	0.11719	0.02451	-0.00021	5.64943	0.15785	0.02295	-0.00020
2007	6.28362	0.12225	0.02309	-0.00019	5.65390	0.16409	0.02141	-0.00018
2008	6.28362	0.12225	0.02309	-0.00019	5.65390	0.16409	0.02141	-0.00018
2009	6.28362	0.12225	0.02309	-0.00019	5.65390	0.16409	0.02141	-0.00018
2010	6.28362	0.12225	0.02309	-0.00019	5.65390	0.16409	0.02141	-0.00018
2011	6.28362	0.12225	0.02309	-0.00019	5.65390	0.16409	0.02141	-0.00018
2012	6.28362	0.12225	0.02309	-0.00019	5.65390	0.16409	0.02141	-0.00018
2013	6.28362	0.12225	0.02309	-0.00019	5.65390	0.16409	0.02141	-0.00018
2014	6.28362	0.12225	0.02309	-0.00019	5.65390	0.16409	0.02141	-0.00018
2015	6.28362	0.12225	0.02309	-0.00019	5.65390	0.16409	0.02141	-0.00018
2016	6.28362	0.12225	0.02309	-0.00019	5.65390	0.16409	0.02141	-0.00018
2017	6.28362	0.12225	0.02309	-0.00019	5.65390	0.16409	0.02141	-0.00018
2018	6.28362	0.12225	0.02309	-0.00019	5.65390	0.16409	0.02141	-0.00018
2019	6.28362	0.12225	0.02309	-0.00019	5.65390	0.16409	0.02141	-0.00018
2020	6.28362	0.12225	0.02309	-0.00019	5.65390	0.16409	0.02141	-0.00018

Table B.5 Summary statistics: CHNS samples

year	**variables**	**urban**				**rural**			
		male		**female**		**male**		**female**	
		mean	**s.d.**	**mean**	**s.d.**	**mean**	**s.d.**	**mean**	**s.d.**
1989	inc.	1408.10	3241.52	1189.11	2372.01	1348.32	3851.84	901.35	1362.31
	e	9.49	3.59	9.20	3.44	8.05	3.17	7.42	3.36
	exp	17.39	11.23	14.58	10.00	16.34	11.01	14.07	9.64
	exp^2	428.29	476.81	312.50	368.49	387.96	440.00	290.84	340.20
1991	inc.	1153.55	808.71	967.68	783.45	900.19	1175.17	744.35	833.47
	e	8.87	4.36	8.16	4.49	7.63	3.59	6.27	4.05
	exp	19.21	11.29	16.99	10.06	18.66	11.44	16.82	10.09
	exp^2	496.32	497.28	389.87	385.10	479.12	487.08	384.78	380.04

Continue to the next page

1993	inc.	2078.45	14505.79	1217.92	1427.18	1582.89	8583.10	959.17	2029.63
	e	10.08	2.72	9.92	2.52	8.99	2.37	8.73	2.25
	exp	19.03	10.29	16.27	8.91	17.96	10.80	15.53	9.33
	exp^2	467.69	446.90	344.01	326.41	439.13	448.84	328.27	334.82
1997	inc.	1893.67	1697.23	1424.90	1119.92	1603.38	2395.21	1603.38	2395.21
	e	10.14	3.23	9.89	3.11	8.75	2.68	8.75	2.68
	exp	20.45	10.78	17.91	9.83	19.69	11.10	19.69	11.10
	exp^2	534.09	465.68	417.36	384.70	510.94	481.11	510.94	481.11
2000	inc.	2222.72	3049.41	1754.13	2286.47	1782.89	2451.73	1412.02	2112.49
	e	10.67	3.16	10.48	3.07	8.96	2.51	8.43	2.60
	exp	21.70	10.86	19.16	9.80	20.77	11.60	18.71	10.06
	exp^2	588.50	477.26	463.01	376.98	565.89	515.33	451.12	400.88

Table B.6 Estimates of the earnings equation: CHNS samples

urban								
year	male				female			
	α	β	γ	δ	α	β	γ	δ
1989	5.35231	0.05317	0.08802	-0.00164	5.23546	0.04431	0.10510	-0.00221
1991	5.38600	0.06477	0.08016	-0.00164	5.28786	0.05821	0.09337	-0.00217
1993	5.55869	0.03794	0.09307	-0.00178	5.03143	0.09546	0.09077	-0.00215
1997	5.69877	0.07534	0.07415	-0.00150	5.37693	0.07968	0.08416	-0.00182
2000	5.60947	0.09419	0.05005	-0.00073	5.52488	0.10861	0.04354	-0.00086
rural								
year	male				female			
	α	β	γ	δ	α	β	γ	δ
1989	5.50869	0.04460	0.06652	-0.00122	5.40414	0.02650	0.08193	-0.00177
1991	4.50197	0.09802	0.08308	-0.00143	4.71452	0.05462	0.11713	-0.00242
1993	4.55075	0.06913	0.12595	-0.00258	4.72135	0.06616	0.09862	-0.00190
1997	4.86154	0.11040	0.08449	-0.00159	5.21502	0.09536	0.05349	-0.00104
2000	5.36047	0.10431	0.06387	-0.00128	5.65727	0.09195	0.02870	-0.00047
ratio								
year	male				female			
	α	β	γ	δ	α	β	γ	δ
1989	0.97161	1.19215	1.32321	1.34426	0.96879	1.67208	1.28280	1.24859
1991	1.19636	0.66078	0.96485	1.14685	1.12161	1.06573	0.79715	0.89669
1993	1.22149	0.54882	0.73894	0.68992	1.06568	1.44287	0.92040	1.13158
1997	1.17222	0.68243	0.87762	0.94340	1.03105	0.83557	1.57338	1.75000
2000	1.04645	0.90298	0.78362	0.56709	0.97660	1.18119	1.51707	1.82049

Table B.7 Fitted ratios of urban-to-rural parameter estimates, CHNS samples

year	male				female			
	α	β	γ	δ	α	β	γ	δ
1985	1.09803	0.83932	1.26307	1.58516	1.06655	1.62640	0.80399	0.80879
1986	1.10016	0.83102	1.21894	1.48731	1.06256	1.57306	0.83867	0.85417
1987	1.10229	0.82281	1.17635	1.39550	1.05859	1.52147	0.87484	0.90209
1988	1.10443	0.81468	1.13525	1.30936	1.05463	1.47157	0.91257	0.95270
1989	1.10658	0.80663	1.09559	1.22853	1.05069	1.42330	0.95193	1.00615
1990	1.10873	0.79866	1.05731	1.15270	1.04676	1.37662	0.99299	1.06260
1991	1.11088	0.79077	1.02037	1.08154	1.04284	1.33147	1.03582	1.12221
1992	1.11304	0.78296	0.98472	1.01478	1.03894	1.28781	1.08050	1.18517
1993	1.11520	0.77522	0.95032	0.95214	1.03506	1.24557	1.12710	1.25167
1994	1.11736	0.76756	0.91711	0.89336	1.03119	1.20472	1.17572	1.32189
1995	1.11953	0.75997	0.88507	0.83822	1.02733	1.16521	1.22643	1.39605
1996	1.12171	0.75246	0.85415	0.78647	1.02349	1.12699	1.27933	1.47437
1997	1.12389	0.74503	0.82431	0.73793	1.01966	1.09003	1.33451	1.55709
1998	1.12607	0.73767	0.79551	0.69238	1.01585	1.05428	1.39207	1.64445
1999	1.12825	0.73038	0.76771	0.64964	1.01205	1.01970	1.45211	1.73671
2000	1.13044	0.72316	0.74089	0.60953	1.00827	0.98626	1.51474	1.83415
2001	1.13264	0.71601	0.71501	0.57191	1.00450	0.95391	1.58008	1.93705
2002	1.13484	0.70894	0.69003	0.53661	1.00074	0.92263	1.64823	2.04572
2003	1.13704	0.70193	0.66592	0.50348	0.99700	0.89237	1.71932	2.16050
2004	1.13925	0.69500	0.64265	0.47240	0.99327	0.86310	1.79348	2.28171
2005	1.14146	0.68813	0.62020	0.44324	0.98956	0.83480	1.87084	2.40972
2006	1.14368	0.68133	0.59853	0.41588	0.98586	0.80742	1.95153	2.54492
2007	1.14590	0.67460	0.57762	0.39021	0.98217	0.78094	2.03570	2.68769
2008	1.14590	0.67460	0.57762	0.39021	0.98217	0.78094	2.03570	2.68769
2009	1.14590	0.67460	0.57762	0.39021	0.98217	0.78094	2.03570	2.68769
2010	1.14590	0.67460	0.57762	0.39021	0.98217	0.78094	2.03570	2.68769
2011	1.14590	0.67460	0.57762	0.39021	0.98217	0.78094	2.03570	2.68769
2012	1.14590	0.67460	0.57762	0.39021	0.98217	0.78094	2.03570	2.68769
2013	1.14590	0.67460	0.57762	0.39021	0.98217	0.78094	2.03570	2.68769
2014	1.14590	0.67460	0.57762	0.39021	0.98217	0.78094	2.03570	2.68769
2015	1.14590	0.67460	0.57762	0.39021	0.98217	0.78094	2.03570	2.68769
2016	1.14590	0.67460	0.57762	0.39021	0.98217	0.78094	2.03570	2.68769
2017	1.14590	0.67460	0.57762	0.39021	0.98217	0.78094	2.03570	2.68769
2018	1.14590	0.67460	0.57762	0.39021	0.98217	0.78094	2.03570	2.68769
2019	1.14590	0.67460	0.57762	0.39021	0.98217	0.78094	2.03570	2.68769
2020	1.14590	0.67460	0.57762	0.39021	0.98217	0.78094	2.03570	2.68769

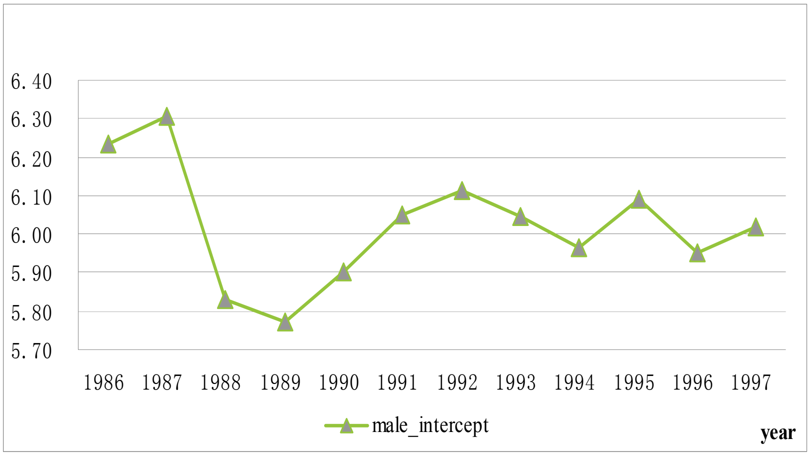
Table B.8 Imputed earnings equation parameters for the rural population, 1985 to 2020

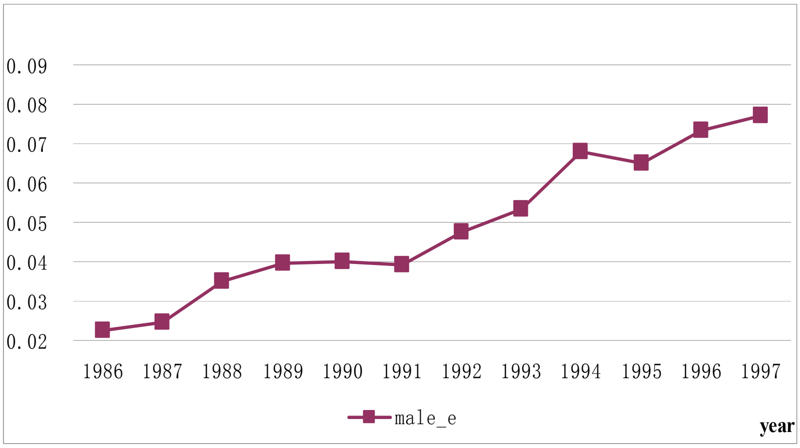
year	male				female			
	α	β	γ	δ	α	β	γ	δ
1985	5.29358	0.01297	0.06773	-0.00093	5.20888	0.01646	0.12262	-0.00258
1986	5.30279	0.01919	0.06613	-0.00090	5.23264	0.02099	0.10967	-0.00219
1987	5.31194	0.02554	0.06456	-0.00088	5.25651	0.02580	0.09809	-0.00186
1988	5.32103	0.03201	0.06303	-0.00085	5.28047	0.03092	0.08773	-0.00157
1989	5.33007	0.03860	0.06154	-0.00083	5.30455	0.03635	0.07846	-0.00133
1990	5.33906	0.04532	0.06008	-0.00080	5.32873	0.04212	0.07017	-0.00113
1991	5.34799	0.05218	0.05866	-0.00078	5.35302	0.04823	0.06276	-0.00096
1992	5.35687	0.05916	0.05727	-0.00076	5.37741	0.05472	0.05613	-0.00081
1993	5.36569	0.06628	0.05591	-0.00074	5.40191	0.06158	0.05020	-0.00069
1994	5.37446	0.07354	0.05459	-0.00071	5.42653	0.06885	0.04490	-0.00058
1995	5.38317	0.08094	0.05330	-0.00069	5.45125	0.07654	0.04016	-0.00049
1996	5.39183	0.08847	0.05204	-0.00067	5.47607	0.08468	0.03592	-0.00042
1997	5.40043	0.09615	0.05080	-0.00066	5.50101	0.09327	0.03212	-0.00035
1998	5.40899	0.10397	0.04960	-0.00064	5.52606	0.10236	0.02873	-0.00030
1999	5.41748	0.11194	0.04843	-0.00062	5.55122	0.11195	0.02569	-0.00025
2000	5.42593	0.12005	0.04728	-0.00060	5.57649	0.12207	0.02298	-0.00022
2001	5.43432	0.12832	0.04616	-0.00058	5.60187	0.13276	0.02055	-0.00018
2002	5.44266	0.13674	0.04507	-0.00057	5.62736	0.14402	0.01838	-0.00015
2003	5.45095	0.14532	0.04400	-0.00055	5.65297	0.15590	0.01644	-0.00013
2004	5.45918	0.15405	0.04296	-0.00054	5.67869	0.16842	0.01470	-0.00011
2005	5.46736	0.16295	0.04194	-0.00052	5.70452	0.18161	0.01315	-0.00009
2006	5.47549	0.17200	0.04095	-0.00051	5.73047	0.19549	0.01176	-0.00008
2007	5.48357	0.18122	0.03998	-0.00049	5.75653	0.21012	0.01052	-0.00007
2008	5.48357	0.18122	0.03998	-0.00049	5.75653	0.21012	0.01052	-0.00007
2009	5.48357	0.18122	0.03998	-0.00049	5.75653	0.21012	0.01052	-0.00007
2010	5.48357	0.18122	0.03998	-0.00049	5.75653	0.21012	0.01052	-0.00007
2011	5.48357	0.18122	0.03998	-0.00049	5.75653	0.21012	0.01052	-0.00007
2012	5.48357	0.18122	0.03998	-0.00049	5.75653	0.21012	0.01052	-0.00007
2013	5.48357	0.18122	0.03998	-0.00049	5.75653	0.21012	0.01052	-0.00007
2014	5.48357	0.18122	0.03998	-0.00049	5.75653	0.21012	0.01052	-0.00007
2015	5.48357	0.18122	0.03998	-0.00049	5.75653	0.21012	0.01052	-0.00007
2016	5.48357	0.18122	0.03998	-0.00049	5.75653	0.21012	0.01052	-0.00007
2017	5.48357	0.18122	0.03998	-0.00049	5.75653	0.21012	0.01052	-0.00007
2018	5.48357	0.18122	0.03998	-0.00049	5.75653	0.21012	0.01052	-0.00007
2019	5.48357	0.18122	0.03998	-0.00049	5.75653	0.21012	0.01052	-0.00007
2020	5.48357	0.18122	0.03998	-0.00049	5.75653	0.21012	0.01052	-0.00007

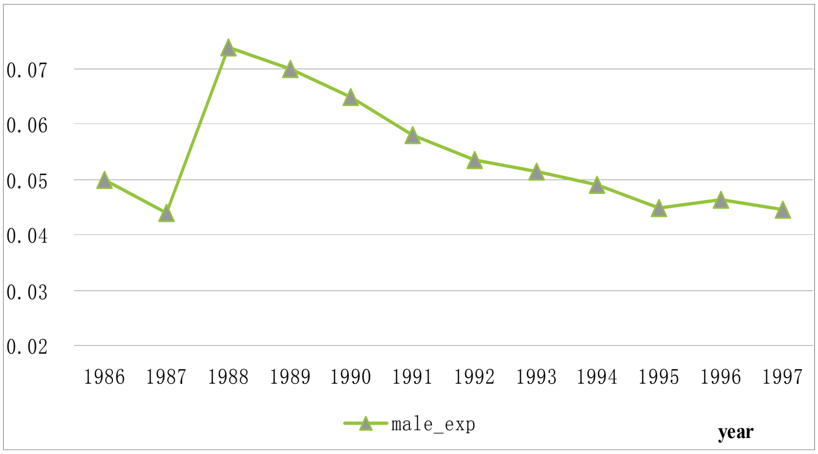
Table B.9 α_ values

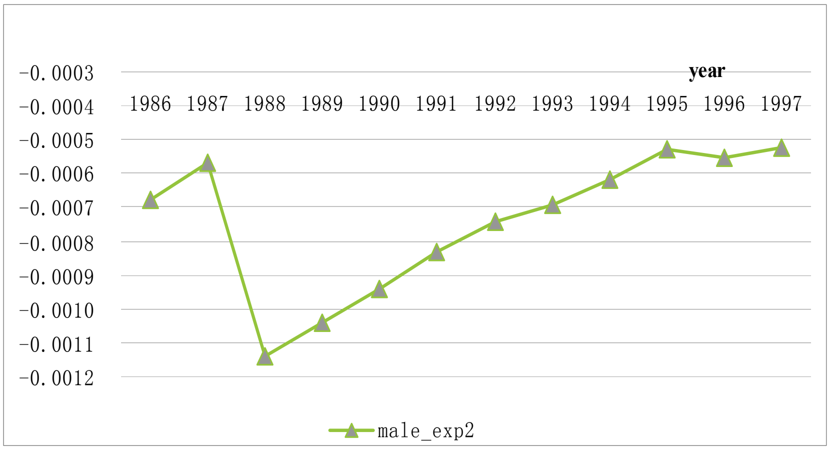
α_value	male	female
1985	1.05637	1.07552
1986	1.05637	1.07552
1987	1.06643	1.09893
1988	1.06640	1.08521
1989	1.06135	1.08638
1990	1.05763	1.07798
1991	1.06023	1.07839
1992	1.07039	1.09548
1993	1.11611	1.13272
1994	1.14799	1.17606
1995	1.13978	1.16159
1996	1.15789	1.19970
1997	1.18900	1.23888
1998	1.18900	1.23888
1999	1.18900	1.23888
2000	1.18900	1.23888
2001	1.18900	1.23888
2002	1.18900	1.23888
2003	1.18900	1.23888
2004	1.18900	1.23888
2005	1.18900	1.23888
2006	1.18900	1.23888
2007	1.18900	1.23888
2008	1.18900	1.23888
2009	1.18900	1.23888
2010	1.18900	1.23888
2011	1.18900	1.23888
2012	1.18900	1.23888
2013	1.18900	1.23888
2014	1.18900	1.23888
2015	1.18900	1.23888
2016	1.18900	1.23888
2017	1.18900	1.23888
2018	1.18900	1.23888
2019	1.18900	1.23888
2020	1.18900	1.23888

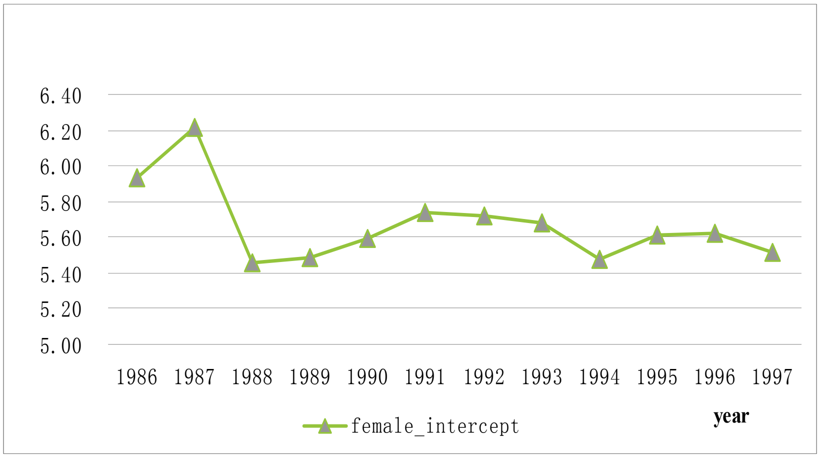
Figures B.1~8 Plotting parameter estimates against time: urban sample

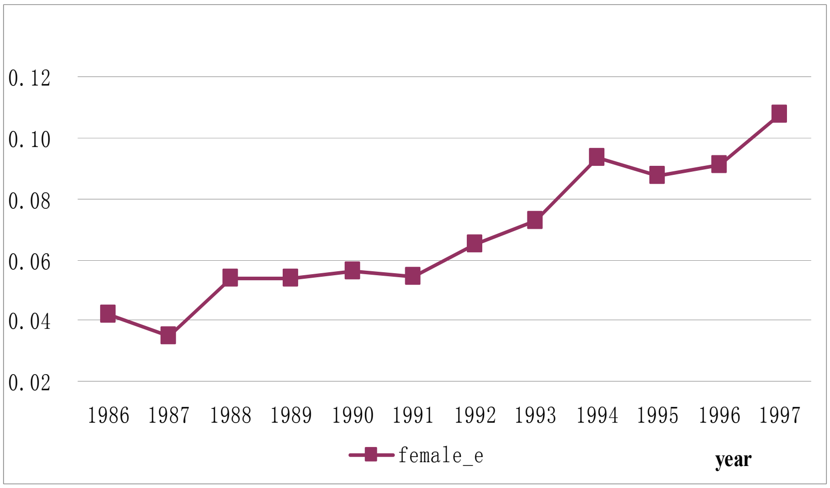


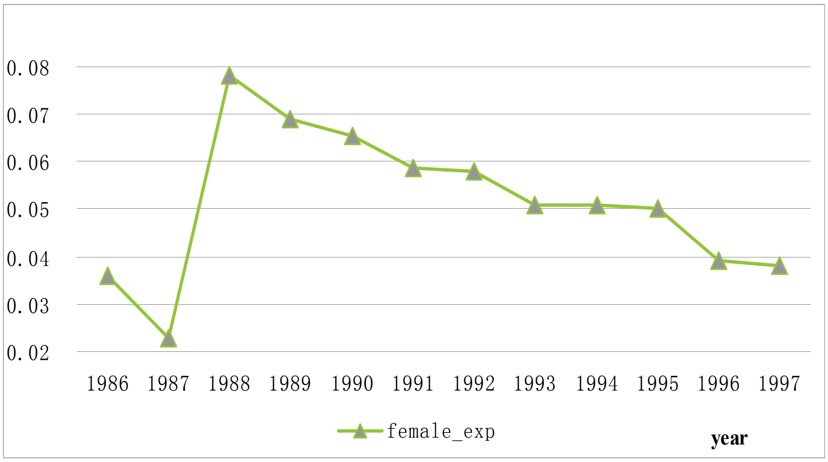


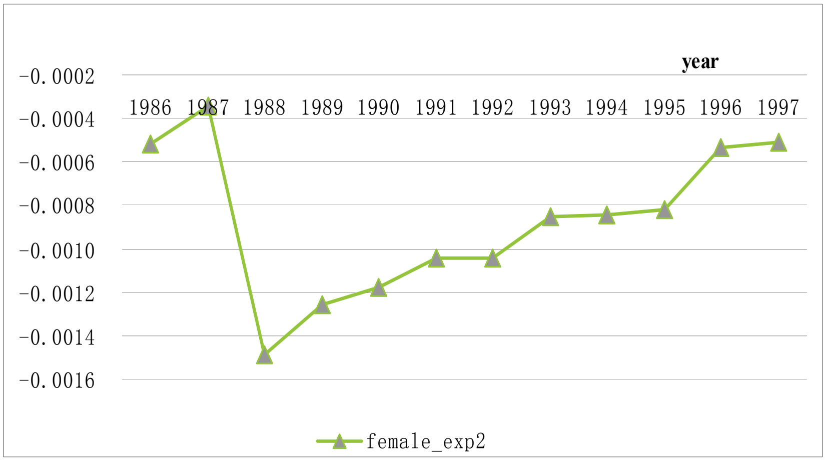



year
-0.0003
-0.0004
-0.0005
-0.0006
-0.0007
-0.0008
-0.0009
-0.0010
-0.0011
-0.0012
1986 1987 1988 1989 1990 1991 1992 1993 1994 1995 1996 1997
male_exp2

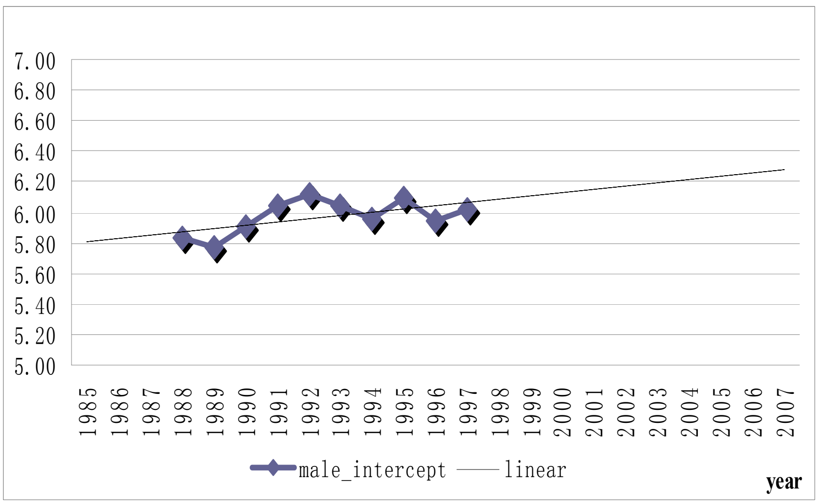

6.40
6.20
6.00
5.80
5.60
5.40
5.20
5.00
1986 1987 1988 1989 1990 1991 1992 1993 1994 1995 1996 1997
female_intercept
year

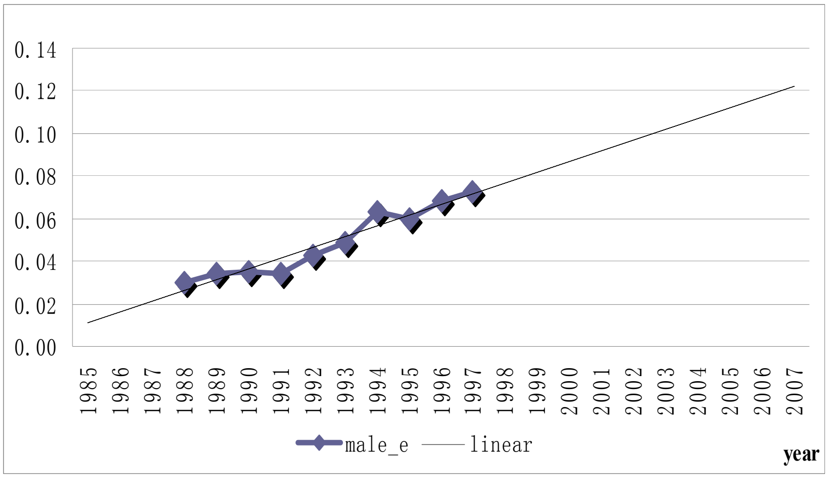

0.12
0.10
0.08
0.06
0.04
0.02
1986 1987 1988 1989 1990 1991 1992 1993 1994 1995 1996 1997
female_e
year

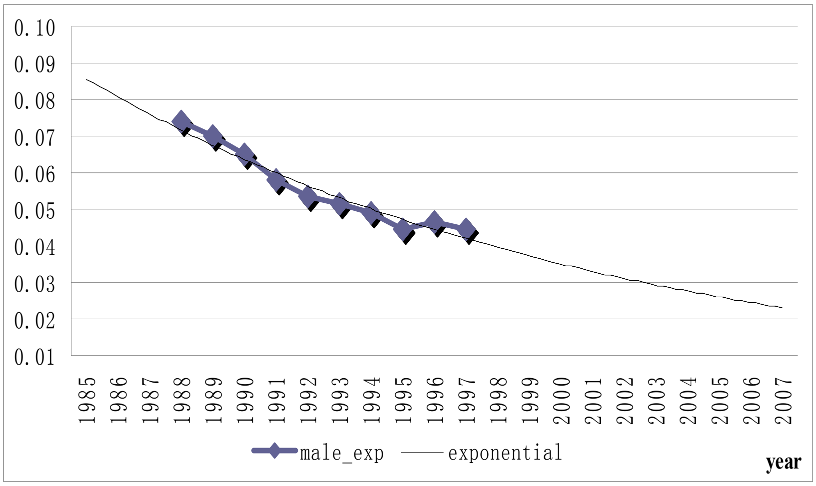


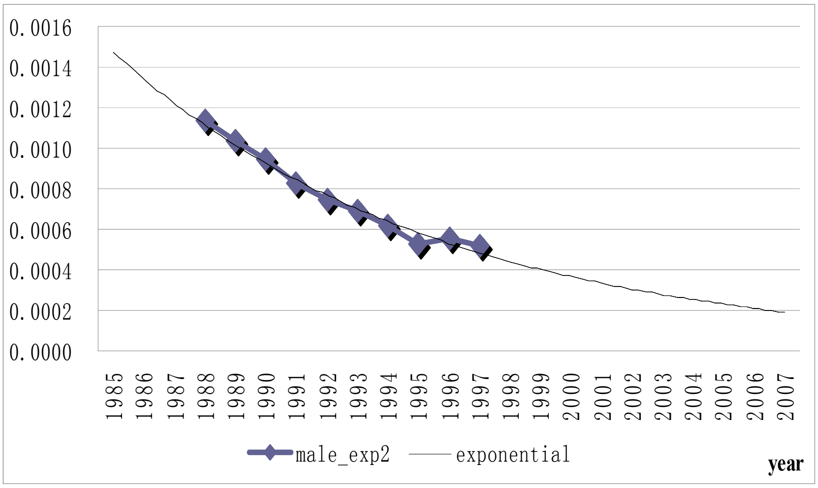


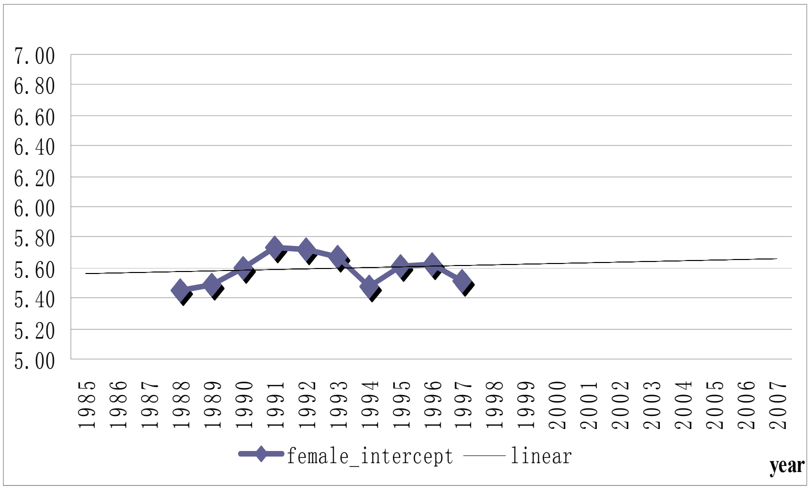
Figures B.9~16 Sample regression lines of parameter estimates, urban sample

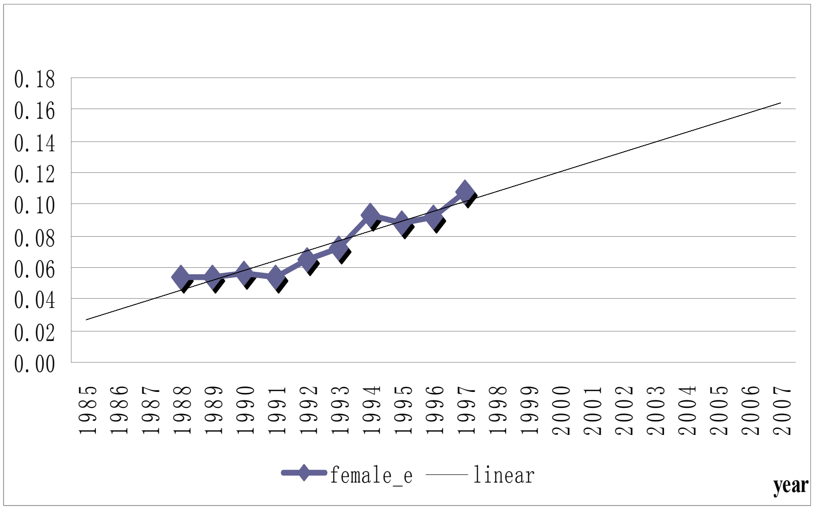


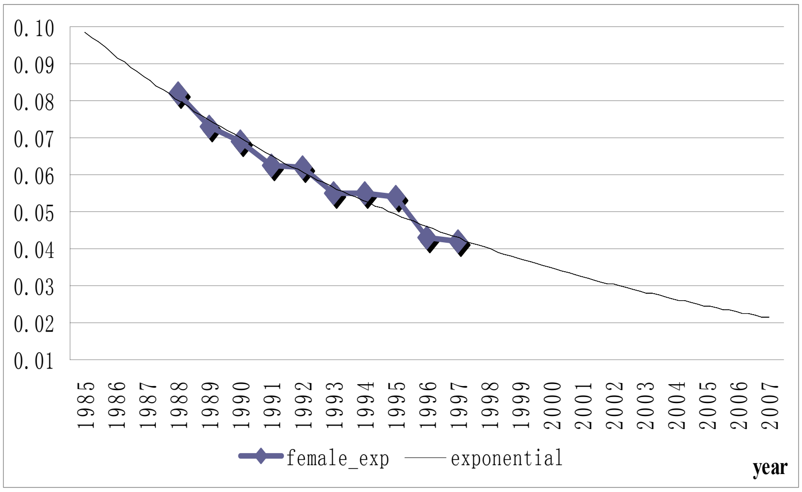

0.14
0.12
0.10
0.08
0.06
0.04
0.02
0.00
1985 1986 1987 1988 1989 1990 1991 1992 1993 1994 1995 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007
male_e
linear
year

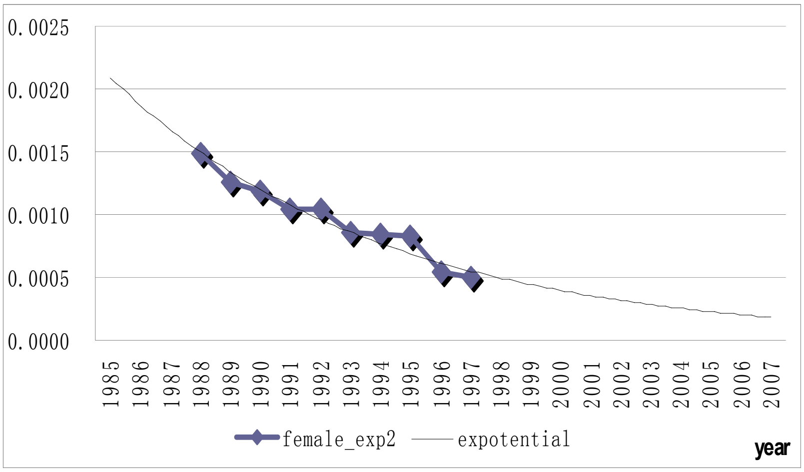

0.10
0.09
0.08
0.07
0.06
0.05
0.04
0.03
0.02
0.01
1985 1986 1987 1988 1989 1990 1991 1992 1993 1994 1995 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007
male_exp
exponential
year


0.0016
0.0014
0.0012
0.0010
0.0008
0.0006
0.0004
0.0002
0.0000
1985 1986 1987 1988 1989 1990 1991 1992 1993 1994 1995 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007
male_exp2
exponential
year

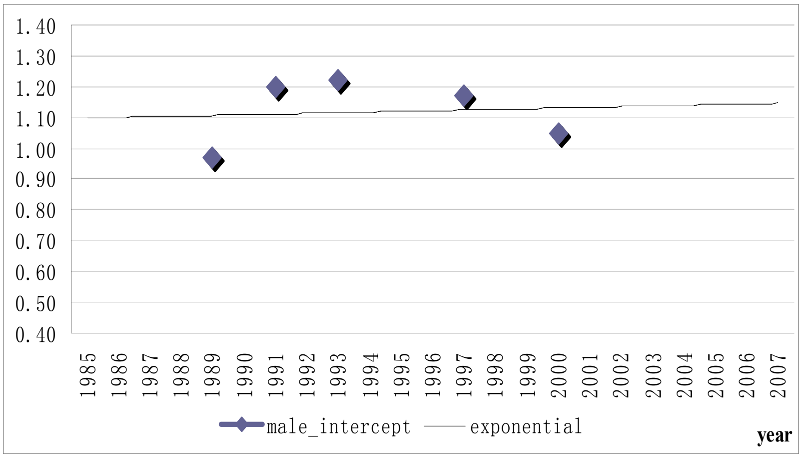

7.00
6.80
6.60
6.40
6.20
6.00
5.80
5.60
5.40
5.20
5.00
1985 1986 1987 1988 1989 1990 1991 1992 1993 1994 1995 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007
female_intercept
linear
year

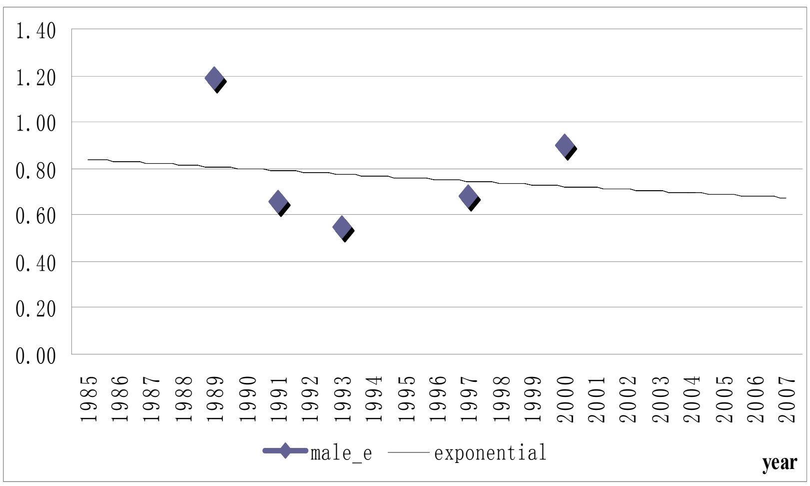

0.18
0.16
0.14
0.12
0.10
0.08
0.06
0.04
0.02
0.00
1985 1986 1987 1988 1989 1990 1991 1992 1993 1994 1995 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007
female_e
linear
year

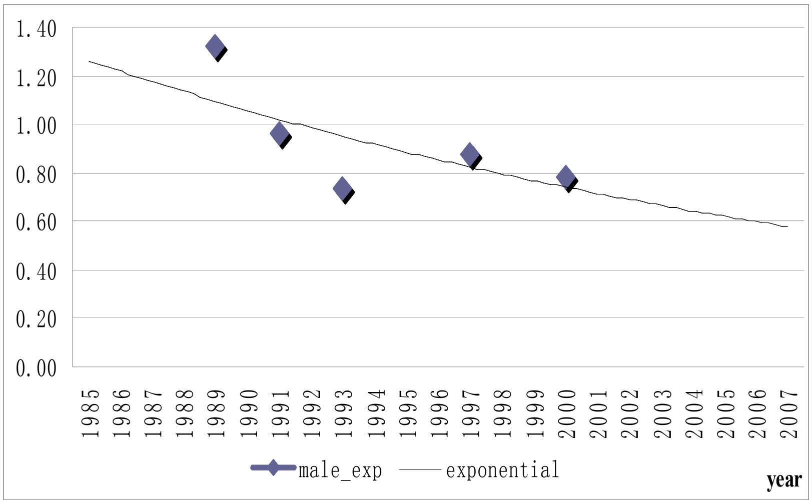

0.10
0.09
0.08
0.07
0.06
0.05
0.04
0.03
0.02
0.01
1985 1986 1987 1988 1989 1990 1991 1992 1993 1994 1995 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007
female_exp
exponential
year

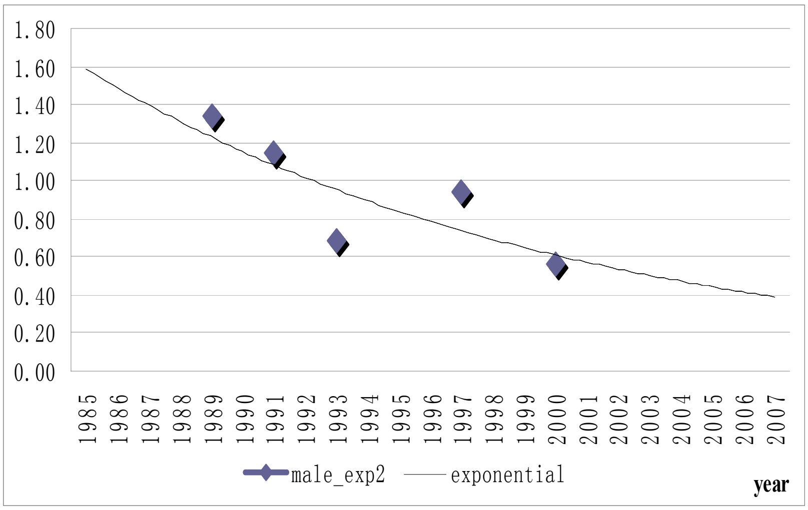


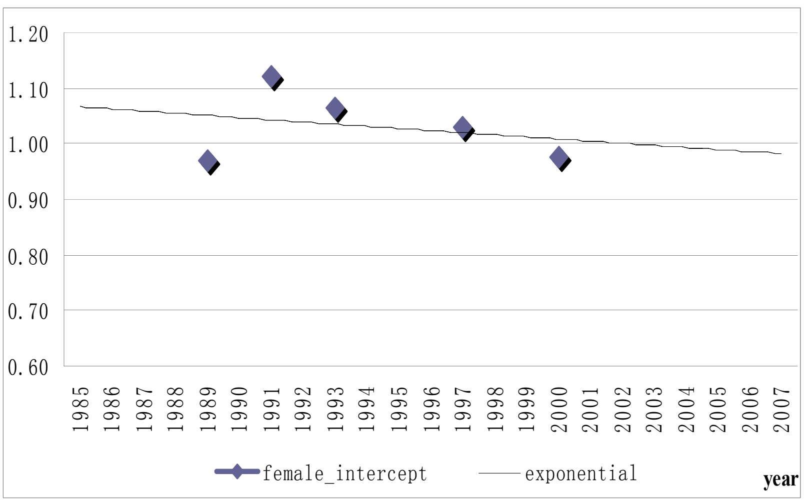
Figures B.17~24: Sample regression lines of urban/rural ratio: CHNS samples

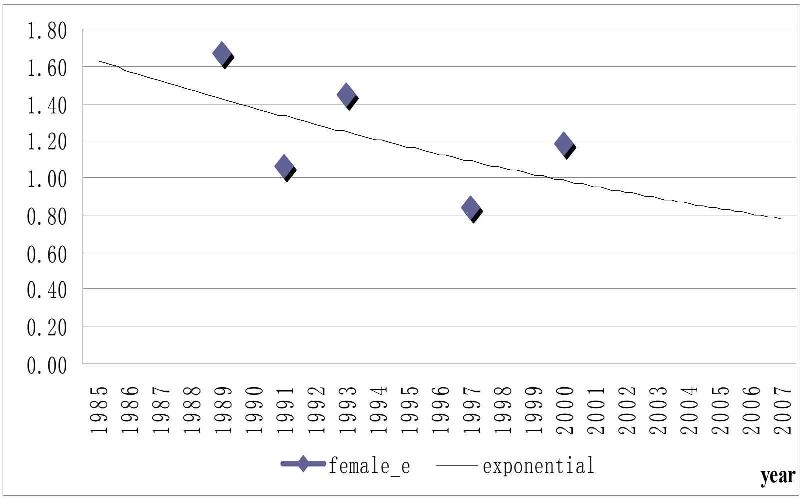


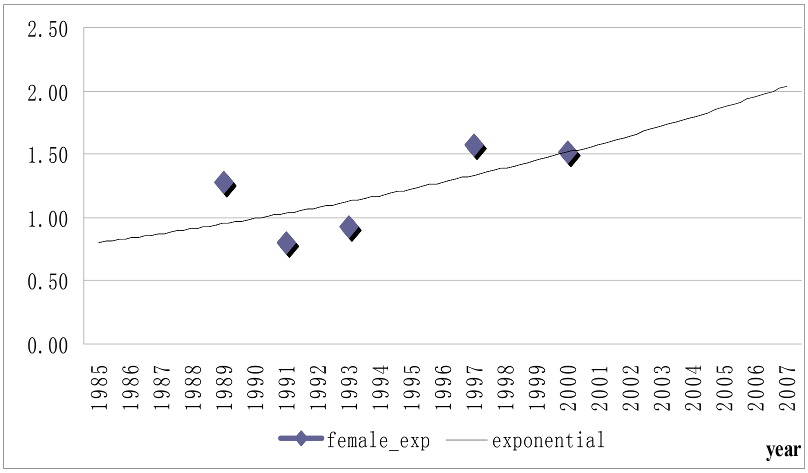


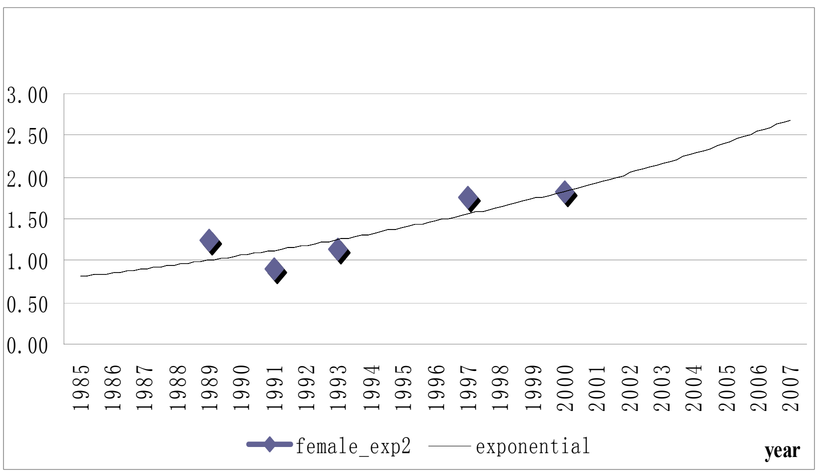

1.40
1.20
1.00
0.80
0.60
0.40
0.20
0.00
1985 1986 1987 1988 1989 1990 1991 1992 1993 1994 1995 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007
male_exp
exponential
year


1.80
1.60
1.40
1.20
1.00
0.80
0.60
0.40
0.20
0.00
1985 1986 1987 1988 1989 1990 1991 1992 1993 1994 1995 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007
male_exp2
exponential
year


1.20
1.10
1.00
0.90
0.80
0.70
0.60
1985 1986 1987 1988 1989 1990 1991 1992 1993 1994 1995 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007
female_intercept
exponential
year


1.80
1.60
1.40
1.20
1.00
0.80
0.60
0.40
0.20
0.00
1985 1986 1987 1988 1989 1990 1991 1992 1993 1994 1995 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007
female_e
exponential
year


2.50
2.00
1.50
1.00
0.50
0.00
1985 1986 1987 1988 1989 1990 1991 1992 1993 1994 1995 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007
female_exp
exponential
year


3.00
2.50
2.00
1.50
1.00
0.50
0.00
1985 1986 1987 1988 1989 1990 1991 1992 1993 1994 1995 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007
female_exp2
exponential
year

Appendix C Human Capital Stock Calculation

This section summarizes the basic methods and procedures of measuring China human capital stock from 1985 to 2020, which are based on the J-F approach. In particular, it explicitly explains the necessary data estimation of the J-F approach based on China's data. We are going to use the following notations:

y = 1980, 1981, 1982, … , 2020, calendar year

s = 1,2, sex, male or female

a = 0, 1, …. 60, age

e: education level, there are two kinds of education categories.

For years 1985-2020 it is classified into five categories: no schooling (ns), primary school (pri), junior middle school (jm), senior middle school (sm), and college (col); For years 2000-2020 it is classified into six categories: no schooling (ns), primary school (pri), junior middle school (jm), senior middle school (sm), college (col) and university (uni). Variables used in measuring the human capital stock are:

- whrs(y,s,a,e): annual market hours worked per person employed, in year y, with sex s, age a , and education level e
- empr(y,s,a,e): employment rate, in year y, with sex s, age a , and education level e
- mhrs(y,s,a,e): market labor time per capita, in year y, with sex s, age a , and education level e
- com(y,s,a,e): hourly compensation, net of taxes on labor income
- yinc(y,s,a,e): annual income of the group employed in year y, with sex s, age a , and education level e
- ymi(y,s,a,e): annual market income per capita, net of tax on labor compensation, in year y, with sex s, age a , and education

level e

- employed(y,s,a,e): the population employed in year y, with sex s, age a , and education level e
- pop(y,s,a,e): the population in year y, with sex s, age a , and education level e
- newEnroll(y,s,a,e): the population enrolled in education level e, in year y, with sex s, age a
- pop_inschool(y,s,a,e-n): the number of the group in school in year y, with sex s, age a, education level e, in grade n+1.
- senr(y,s,a,e+1,e-n): the possibility of the group enrolled in the next education level e+1 in school in year y, with sex s, age a , education level e, and grade n+1.
- mi(y,s,a,e): the lifetime income of the group not in school in year y, with sex s, age a , and education level e.
- R = (1+real growth rate of income)/(1+discount rate)
- pop_inschool(y,s,a,e): the number of the group in school in year y, with sex s, age a , and education level e.
- pop_nischool(y,s,a,e): the number of the group not in school in year y, with sex s, age a , and education level e.
- Le(y)：total population with education level e in year y.
- Ls(y)：total population with sex s.
- Mi(s): the lifetime income for both sexes (nominal income).
- v_e： the shares of current value lifetime income for population with education level e.
- $\overline{v}_e$： the average shares of the current value lifetime income for population with education level e.
- $\overline{v}_s$： the average shares of the current value lifetime income for population with sex s.

- ΔlnK：the growth rate of aggregate human capital stock
- Poplog(y,s): the logarithmic growth rate of the population for both sexes in year y.
- Mitg (y): cumulated the growth rate of aggregate human capital stock
- MiQ(y): the quantity total lifetime income of the country in year y according to the base year.

1 Age categories for calculating lifetime income in the J-F approach

no school or work	0-5
school only	6-16
work and school	16- a
work only	a -59
retirement	male：60+；female：55+

(1) When calculating the lifetime income based on the J-F approach, the retirement age is 60 for male and 55 for female, according to China's law. The legal retirement age is the retirement age approved principally in the second meeting of the fifth NPC Standing Committee on May 24, 1978, which is still in effect and regulated by "The Temporary Method of Settling the Old, Weak, Ill, and Disabled Cadre by the State Council" and "The Temporary Method of the Retired Workers by the State Council" (1978, No.104). The legal retirement age is 60 for male, 50 for female workers and 55 for female cadres. Working in, high temperature, high elevation, highly exhausting conditions, harmful conditions, the legal retirement age is 55 for male and 45 for female. Losing ability to work completely, due to illness and disability not caused by the work, and approved by the

hospital and labor appraisal committee, the legal retirement age is 50 for male and 45 for female.

(2) The *a* in the table is the upper bound of "work and school", and the lower bound of "work only". This age is regulated according to the calculation of the lower bound of people in school in each year. The method of calculating the people in school is discussed in section 3.2.

2 The estimation of annual market income ymi(y,s,a,e)

2.1 The estimation of annual income of the employed

2.1.1 The estimation of annual income of the employed using Mincer equation.

Using the survey data of CHIP (Chinese Household Income Project), CHNS (China Health and Nutrition Survey), and UHS (Urban Household Survey), we regress the logarithm of yearly income *ln yinc* on years of schooling *s*, work experience *exp* and the square of work experience exp^2 by OLS.

$$\ln \text{yinc} = \alpha + \beta e + \gamma \exp + \delta \exp^2 + \text{u}$$

We use the fitted value of *ln yinc* from the equation above to obtain $\text{m}_\text{i} = e^{\ln \text{yinc}}$. Then we use the yearly income observed in the survey data as the dependent variable, regress it on the independent variable m_i by OLS (without the intercept), we obtain the parameter α [23]. Finally the calculate yearly income of the employed as yinc=$\alpha \times e^{\ln \text{yinc}}$.

Note: The yearly income estimated by Mincer equation is the real yearly income (1985 as based year).

[23] Jeffrey M. Wooldridge (2005), Introductory Econometrics: A Modern Approach, 3rd edition.

2.1.2 The years of schooling and working experience in the Mincer equation

(1) Years of schooling:

	No schooling	Primary school	Junior middle school	Senior middle school	College	University
1985-1999	0	6	9	12	15	
2000-2007	0	6	9	12	15	16

(2) Work experience:

For people younger than16, working experience is 0: *exp*=0;

For people older than16, if s<10, working experience: *exp*=age -6;

For people older than16, if s≥10, working experience: *exp*=age-schooling-6.

2.2 The estimation of annual market income

When estimate the yearly income of the employed using the Mincer equation, we obtain $yinc_{y,s,a,e} = whrs_{y,s,a,e} \times com_{y,s,a,e}$.

According to

$mhrs_{y,s,a,e}=whrs_{y,s,a,e} \times empr_{y,s,a,e}$, $ymi_{y,s,a,e}=whrs_{y,s,a,e} \times empr_{y,s,a,e} \times com_{y,s,a,e}$,

we can translate the formula in the J-F approach to:

$ymi_{y,s,a,e}=yinc_{y,s,a,e} \times empr_{y,s,a,e}$.

2.2.1 The calculation of employment rate empr(y,s,a,e)

For the employment rate empr(y,s,a,e) by age, sex and education level for people older than 16, we use the average of the employment rates of 1995 and 2000. The people employed and the total population in 1995 only include the education level of college and above, so we assume that the employment rate is the same for people at this education level.

The formula to calculate the employment rate is:

empr(y,s,a,e)=[employed(y,s, a, e)]/pop(y,s, a, e)

The data sources of employment rate are listed in the table below:

Data	Sources
The employed by age, sex, and education level in 1995	"China Population Statistical Yearbook 2000"
Population by age, sex, and education level in 1995	"China Population Statistical Yearbook 1999"
The employed by age, sex, and education level in 2000	"China Population Census 2000"
Population by age, sex, and education level in 2000	"China Population Census 2000"

Note: The 1% sample population in 1995 is converted into the whole population by the actual sampling percentage of 1.04%.

The employed in "China Population Census 2000" of each province, autonomous region and municipality directly under the central government is converted into the whole population employed by the actual sampling percentage of 10%.

3 Calculation of enrollment rate

Enrollment rate is the probability that a group with education level *e* is enrolled in a higher education level *e*+1.

3.1 Calculation of enrollment by sex, age and education level

According to the age distribution of enrollment number for a certain education level and sex, the formula for the enrollment number of each year by sex, age and education level is:

NewEnroll(y,s,a,e)=NewEnroll(y,s,e)*[NewEnroll(y,s,a,e)/ NewEnroll (y,s,e)]

Note: [NewEnroll (y,s,a,e)/ NewEnroll (y,s,e)] refers to the age distribution of enrollment number for each education level and sex, which is consistent with the age distribution when estimating the population using the perpetual inventory method.

There is no college or university in rural area, so the enrollment number of college and university in rural areas is assumed to be 0.

3.2 In-school population of each education level and each grade

The in-school population of age *a*, sex *s*, education level *e*, grade *n*+1 in year *y* is the enrollment population of age a-n, sex s , education level e in year y-n:

pop_inschool(y,s,a,e-n) = NewEnroll (y-n, s, a-n, e)

3.3 Enrollment rate of each education level and each grade

The probability of advancing to the next higher education level is estimated as the average ratio of the sum of all students of any age in a year who are initially enrolled to the sum of all students of any age initially enrolled in the next higher education level *X* years later. "*X*" depends upon the number of years it takes to complete an education level.

3.3.1 Enrollment rate from no schooling to primary school

The formula from no schooling to primary school is:

senr(y,s,a,pri-ns) = Newenroll(y+1,s,a+1,pri)/pop(y,s,a,ns)

The upper bound of people out of school in year y and enroll into primary school in year y+1 is determined by the upper bound of age distribution for enrollment of primary school in year y+1. For example, the age distribution for enrollment of primary school in year y+1 is from 6 to 12, the upper bound of people no schooling in year y and enrolled into primary school in year y+1 is 11. The upper bound of people out of school in 2007 and enroll into primary school in 2008 is the same for 2006.

3.3.2 Enrollment rate from primary school to junior middle school

The steps of calculating this enrollment rate by sex and age in year y are:

(1) The enrollment rate of the first grade of primary school in year y by age and sex is the average enrollment rate that the group in this grade can enroll in the first grade of junior middle school six years later, and the formula is:

senr(y,s,a,jm-pri) = newEnroll (y+6, s, jm)/ newEnroll (y, s, pri)

(2) The population of the second grade of primary school in year y by age and sex is the enrollment population of primary school in year y-1 by age and sex. The probability of the group in this grade can enroll in junior middle school 5 years later is the average enrollment rate that the group in this grade can enroll in the first grade of junior middle school five years later, and the formula is:

senr(y,s,a,jm-pri-1) = newEnroll (y+5, s, jm)/ newEnroll (y-1, s, pri)

(3) The population of the third grade of primary school in year y by age and sex is the enrollment population of primary school in year y-2 by age and sex. The probability of the group in this grade can enroll in junior middle school 4 years later is the average enrollment rate that the group in this grade can enroll in the first grade of junior middle school four years later, and the formula is:

senr(y,s,a,jm-pri-2) = newEnroll (y+4, s, jm)/ newEnroll (y-2, s, pri)

(4) Similarly, we can calculate the probability of the group of each grade in primary school that enroll in junior middle school in year y.

3.3.3 Enrollment rate from junior middle school to senior middle school

The steps of calculating this enrollment rate by sex and age in year *y* are:

(1) The enrollment rate of the first grade of junior middle school in year y by age is the average enrollment rate that the group in this grade can

enroll in the first grade of senior middle school three years later, and the formula is:

senr(y,s,a,sm-jm) = newEnroll (y+3, s, sm)/ newEnroll (y, s, jm)

(2) The population of the second grade of junior middle school in year y by age and sex is the enrollment population of junior school in year y-1 by age and sex. The probability of the group in this grade can enroll in senior middle school two years later is the average enrollment rate that the group in this grade can enroll in the first grade of senior middle school two years later, and the formula is:

senr(y,s,a,sm-jm-1) = newEnroll (y+2, s, sm)/ newEnroll (y-1, s, jm)

(3) Similarly, we can calculate the probability of the group of each grade in junior middle school that enroll in senior middle school in year y.

3.3.4 Enrollment rate from senior middle school to college or university

The steps of calculating the enrollment rate from senior middle school to college by sex and age in year *y* are:

(1) The enrollment rate of the first grade of senior middle school in year *y* by age is the average enrollment rate that the group in this grade can enroll in the first grade of college three years later, and the formula is:

senr(y,s,a,col-sm) =newEnroll (y+3, s, col)/ newEnroll (y, s, sm)

(2) The population of the second grade of senior middle school in year y by age and sex is the enrollment population of senior school in year y-1 by age and sex. The probability of the group in this grade can enroll in college two years later is the average enrollment rate that individuals in this grade can enroll in the first grade of college two years later, and the formula is:

senr(y,s,a,col-sm-1) = newEnroll (y+2, s,col)/ newEnroll (y-1, s, sm)

(3) Similarly, we can calculate the probability of the group of each grade in senior middle school that can enroll in college in year *y*.

The steps of calculating the enrollment rate from senior middle school to university by sex and age in year *y* are:

(1) The enrollment rate of the first grade of senior middle school in year y by age is the average enrollment rate that the group in this grade can enroll in the first grade of university three years later, and the formula is:

senr(y,s,a,col-uni) =newEnroll (y+3, s, uni)/ newEnroll (y, s, sm)

(2) The population of the second grade of senior middle school in year y by age and sex is the enrollment population of senior school in year y-1 by age and sex. The probability of the group in this grade can enroll in university two years later is the average enrollment rate that the group in this grade can enroll in the first grade of university two years later, and the formula is:

senr(y,s,a, uni -sm-1) = newEnroll (y+2, s,uni)/ newEnroll (y-1, s, sm)

(3) Similarly, we can calculate the probability of the group of each grade in senior middle school that can enroll in university in year *y*.

Note: 1) By using different years' enrollment population in the calculation of enrollment rate, an adjustment has already been made for survival rate. Therefore, the survival rate is not included in the formula. We also assume that no one drops out, skips a grade, repeats, or stays out for a year or more within a certain education category.

2) After 2002, fix the enrollment rates from the last available year.

4 Lifetime income calculation for in-school population

The number of years discounted until they realize the higher level of lifetime income depends on the number of years it takes to complete the starting grade level and the current grade of enrollment within the starting grade level.

4.1 Lifetime income of population in primary school by age and sex

(1) If the individual in the first grade of primary school can advance to the next higher education level, he could get lifetime income equal to someone who is currently six years older and whose educational attainment is junior middle school, and discount 6 years before he realizes junior middle school: senr(y,s,a,jm-pri)*mi(y,s,a+6,jm)*R^6.

(2) If the individual in the second grade of primary school can advance to the next higher education level, senr(y,s,a,jm-pri-1)*mi(y,s,a+5,jm)*R^5 is calculated by similar way.

(3) Similarly, we can calculate the lifetime income of the group in each grade of primary school.

4.2 Lifetime income of the group in junior middle school and above by age and sex

Take junior middle school as an example：

(1) If the individual in the first grade of junior middle school can advance to the next higher education level, he could get lifetime income equal to someone who is currently three years older and whose educational attainment is senior middle school, and discount 3 years before he realizes senior middle school: *senr(y,s,a,sm-jm)*mi(y,s,a*+3*,sm)**R^3

(2) If the individual in the second grade of junior middle school can advance to the next higher education level, *senr(y,s,a,sm-jm*-1*)* mi(y,s,a*+2*,sm)**R^2 is calculated by similar way.

(3) Similarly, we can calculate the lifetime income of the group in each grade of junior middle school.

For the years that do not separate university from college (there are five categories for education level, and the last level is college and above), to get the lifetime income of the group in the first grade of senior middle school, we should use the equation $senr(y,s,a,col\text{-}sm)*mi(y,s,a+3,col)*R^3$, $senr(y,s,a,col\text{-}sm\text{-}1)*mi(y,s,a+2,col)*R^2$ for grade 2 students, and so on.

For the years that separate university from college (there are six categories for education level, and the last level is university and above), we should use the equation $senr(y,s,a,col\text{-}sm)*mi(y,s,a+3,col)*R^3+senr(y,s,a,uni\text{-}sm)*mi(y,s,a+3,uni)*R^3$, as for a senior middle school students, they can go to college or university after their graduation. Calculate $senr(y,s,a,col\text{-}sm\text{-}1)*mi(y,s,a+2,col)*R^2+senr(y,s,a,uni\text{-}sm\text{-}1)*mi(y,s,a+2,uni)*R^2$ for grade 2 students, and so on. Similarly, we can calculate the lifetime income of the group in each grade of senior middle school.

Note: By using the average ratio of the sum of all students of any age in a year who are initially enrolled to the sum of all students of any age initially enrolled in the next higher education level "X" years later, an adjustment has already been made for survival. Accordingly there is no survival rate in the formula.

5 Out-of-school population's lifetime income

5.1 Calculation of out of school population

In-school population of age a, sex s, education level e in year y, pop_inschool(y,s,a,e), is the sum of population of each grade:

$$pop\_inschool(y,s,a,e)=\sum_{n=0}^{y(e)} pop\_inschool(y,s,a,e-n)$$

y(e) is the number of years to achieve education level e. The formula for calculating out-of-school population of age a, sex s, education level e in year y is:

pop_nischool(y,s,a,e) = pop(y, s, a,e) - pop_inschool(y,s,a,e)

Note: Following adjustment is necessary for negative values in out-of-school population

(1) Reset negative out-of-school population for certain gender, age, education level to 0. The negative out-of-school population mainly appears in primary school for students aging 5-10.

(2) Add the negative out-of-school population for certain gender, age, education level to the in-school population by grades. The weights to add the negative number are the proportion of certain grade of the total population in school with this gender, age, education level.

5.2 Out-of-school category's lifetime income

The out-of-school population is the population that works only, we can use the fourth and fifth stages of J-F approach to calculate the lifetime income for this category.

When age < 60, the formula of lifetime income is:

$$mi(y,s,a,e) = ymi(y,s,a,e) + sr(y+1,s)*mi((y,s,a+1,e)*R$$

When age > 60, lifetime income is zero, i.e. ymi = 0.

6 The growth rate of real wage and discount rate

6.1 The growth rate of real wage

We use the average labor productivity growth rate to be the real wage growth rate of urban and rural areas. Moreover, we use the labor productivity of the primary sector as the rural labor productivity, and the labor productivity of the secondary and tertiary sectors as the urban labor productivity. The values are 6% for urban area and 4.11% for rural area.

6.2 Discount rate

We use average real interest rate of long-term government bonds, the value for our sample period is 3.14%

7 Tax rate and non-market income

(1) We use Mincer equations to estimate the employed population's annual income. For incomes reported in CHIP, CHNS and UHS, it is not clear whether it is the after-tax income; therefore we do not deduct tax when estimating the average market annual income.
(2) Non-market lifetime income is not included in the calculation; the final human capital stock is derived from market income only.

8 Calculation of Divisia index

8.1 Calculating the Divisia index using educational level information

The steps of calculating the Divisia index using educational level information are as follows:

(1) Using Tornqvist aggregation, the growth rate of aggregate human capital stock is calculated as a weighted sum of the growth rates of the number of individuals across different educational level categories of population:

$$\Delta \ln K^{edu} = \sum_{e} \bar{v}_e \Delta \ln L_e$$

where ΔlnK denotes the growth rate of aggregate human capital stock, L_e denotes the number of individuals with educational level e. Δ denotes a first difference, or change between two consecutive years,

$$\Delta \ln L_e = \ln L_e(y) - \ln L_e(y-1)$$

where y denotes the year.

Weights are given by current dollar lifetime income share for each educational level of the population in the aggregate current dollar lifetime income：

$$\bar{v}_e = \frac{1}{2}[v_e(y) + v_e(y-1)], v_e = \frac{Mi_e}{\sum_e Mi_e}$$

where Mi_e is the total current dollar market lifetime income of individuals with educational level e.

(2) We use year 2001 as the base year

(3) Add up (cumulate) the growth rate of aggregate human capital stock to obtain the cumulated growth rates Mitg(y) for year *y*,

$$Mitg(y) = \sum_{1986}^{y} \Delta \ln K$$

That is,

Mitg(1986) = ΔlnK(1986)

Mitg(1987) = Mitg(1986) + ΔlnK(1987)

……

Mitg(2007) = Mitg(2006) + ΔlnK(2007)

(4) Take the exponential of all of the added up (cumulated) growth rates Mitg(y) for year *y*,

$$Mitg\exp(y) = \exp[Mitg(y)]$$

(5) Then normalize these results, so that the Mi, Mi(Q), in year *y*,

$$MiQ(y) = \frac{Mitg\exp(y) \cdot Mi(b)}{Mitg\exp(b)}$$

Where Mi(b) is the total current dollar lifetime market income in the base year. After normalization, in the base year the quantity of human capital MiQ is equal to current dollar human capital Mi.

8.2 Quantity index using gender information

For indexes based on gender information, the number of weighted growth rates is *s* sub-aggregate components (male, female). Similarly, the growth rate of aggregate human capital stock is calculated as a weighted sum of the growth rates of the number of individuals across different gender of the population

$$\Delta \ln K^{gender} = \sum_s \bar{v}_s \Delta \ln L_s$$

When computing the total growth rates of the human capital stock, we continue the step (2) to (5) as in 8.1, and obtain the index.

9 The human capital stock of year 1985-2020 in China

When calculate the human capital, the income estimated by Mincer equation is the real yearly income (based year is 1985), and then we use CPI to inflate and obtain the nominal yearly income.

Tables C.1~8 give the real human capital in China in 1985-2020. We also create a new human capital series starting from 2000, as the reported education categories separate college and university or above. After 2007, we use the population forecast and all other values as of 2007 to forecast the human capital in China.

Tables and Figures of Appendix C

Table C.1 Real urban total human capital 1985-2020, in trillions

year	real urban total human capital	male	female
1985	10.95	7.12	3.84
1986	11.84	7.67	4.17
1987	12.94	8.29	4.65
1988	13.84	8.98	4.86
1989	14.66	9.48	5.19
1990	15.61	10.12	5.50
1991	16.80	10.99	5.81
1992	18.01	11.71	6.30
1993	19.80	12.83	6.97
1994	21.69	13.88	7.81
1995	22.93	14.62	8.31
1996	26.88	17.06	9.81
1997	31.67	20.04	11.62
1998	35.31	22.36	12.95
1999	40.15	25.15	15.00
2000	44.51	27.71	16.80
2001	48.41	30.11	18.31
2002	53.07	32.79	20.27
2003	58.37	35.79	22.58
2004	62.17	38.40	23.76
2005	66.93	41.57	25.36
2006	72.36	44.55	27.81
2007	78.50	48.43	30.07
2008	79.45	49.08	30.37
2009	80.89	50.00	30.89
2010	82.20	50.78	31.42
2011	83.98	51.90	32.08
2012	85.74	52.97	32.77
2013	87.56	54.09	33.47
2014	89.37	55.19	34.18
2015	91.07	56.23	34.84
2016	92.57	57.16	35.41
2017	94.04	58.09	35.95
2018	95.52	59.02	36.49
2019	96.91	59.89	37.02
2020	98.19	60.69	37.49

Note: The results are for five education categories.

Table C.2 Real urban total human capital 2000-2020, in trillions

year	real urban total human capital	male	female
2000	45.81	28.39	17.42
2001	49.88	30.87	19.01
2002	54.81	33.69	21.12
2003	60.45	36.85	23.60
2004	64.46	39.59	24.88
2005	69.58	42.95	26.63
2006	75.46	46.16	29.31
2007	82.08	50.29	31.79
2008	83.08	50.97	32.10
2009	84.60	51.94	32.66
2010	86.01	52.76	33.24
2011	87.91	53.95	33.96
2012	89.80	55.09	34.71
2013	91.75	56.28	35.47
2014	93.68	57.44	36.24
2015	95.50	58.53	36.97
2016	97.10	59.52	37.58
2017	98.66	60.50	38.16
2018	100.24	61.48	38.75
2019	101.73	62.40	39.33
2020	103.09	63.25	39.84

Note: The results are for six education categories.

Table C.3 Real rural total human capital 1985-2020, in trillions

year	real rural total human capital	male	female
1985	16.03	8.73	7.29
1986	16.19	9.01	7.18
1987	16.44	9.23	7.21
1988	16.76	9.66	7.11
1989	17.02	9.95	7.07
1990	17.41	10.40	7.01
1991	17.84	10.95	6.89
1992	18.46	11.45	7.01
1993	19.68	12.42	7.26
1994	21.04	13.29	7.76
1995	21.68	13.83	7.85
1996	22.88	14.65	8.23
1997	24.35	15.60	8.75

Continue to the next page

1998	25.17	16.24	8.93
1999	26.31	17.00	9.31
2000	27.69	17.93	9.76
2001	28.63	18.43	10.21
2002	29.56	18.84	10.72
2003	30.83	19.43	11.40
2004	32.42	20.17	12.25
2005	34.85	21.37	13.48
2006	37.10	22.19	14.91
2007	40.25	23.58	16.67
2008	39.41	23.11	16.30
2009	38.49	22.59	15.90
2010	37.58	22.07	15.51
2011	36.78	21.61	15.17
2012	35.94	21.14	14.80
2013	35.16	20.71	14.45
2014	34.42	20.30	14.12
2015	33.71	19.91	13.80
2016	33.03	19.54	13.49
2017	32.35	19.18	13.17
2018	31.71	18.84	12.87
2019	31.09	18.51	12.59
2020	30.36	18.10	12.26

Note: The results are for five education categories.

Table C.4 Real rural total human capital 2000-2020, in trillions

year	real rural total human capital	male	female
2000	27.69	17.93	9.76
2001	28.64	18.43	10.21
2002	29.57	18.84	10.72
2003	30.85	19.44	11.41
2004	32.44	20.19	12.25
2005	34.88	21.38	13.49
2006	37.14	22.21	14.93
2007	40.30	23.61	16.69
2008	39.46	23.14	16.33
2009	38.56	22.62	15.93
2010	37.65	22.10	15.55
2011	36.85	21.65	15.21
2012	36.01	21.18	14.84

Continue to the next page

2013	35.24	20.75	14.49
2014	34.51	20.34	14.16
2015	33.80	19.95	13.84
2016	33.12	19.58	13.53
2017	32.45	19.23	13.22
2018	31.81	18.89	12.92
2019	31.19	18.55	12.64
2020	30.46	18.15	12.31

Note: The results are for six education categories.

Table C.5 Real urban average human capital, 1985-2020 unit: yuan

year	real urban average human capital	male	female
1985	47,874	58,718	35,653
1986	49,445	60,867	36,750
1987	51,671	63,392	38,847
1988	53,269	65,544	39,569
1989	54,687	66,565	41,241
1990	56,851	69,018	42,924
1991	59,528	73,340	43,905
1992	62,253	76,554	46,209
1993	66,830	82,689	49,387
1994	71,541	87,563	53,989
1995	73,996	91,024	55,665
1996	81,441	99,423	61,962
1997	90,412	109,776	69,320
1998	95,361	115,128	73,550
1999	102,885	122,988	80,753
2000	108,553	128,636	86,319
2001	113,484	134,902	89,989
2002	119,520	142,030	95,132
2003	126,543	149,527	101,754
2004	131,048	156,205	103,983
2005	137,882	165,300	108,406
2006	146,019	173,136	116,727
2007	154,803	183,536	123,629
2008	153,427	181,627	122,649

Continue to the next page

2009	152,987	180,708	122,552
2010	152,552	179,586	122,701
2011	152,905	179,612	123,259
2012	153,122	179,443	123,775
2013	153,620	179,662	124,463
2014	153,963	179,799	124,968
2015	154,009	179,980	124,921
2016	153,864	179,932	124,700
2017	153,423	179,713	124,091
2018	153,654	179,831	124,374
2019	154,169	179,605	125,434
2020	154,089	178,859	125,872

Note: The results are for five education categories.

Table C.6 Real urban average human capital, 2000-2020 unit: yuan

year	real urban average human capital	male	female
2000	111,730	131,804	89,506
2001	116,929	138,316	93,468
2002	123,440	145,896	99,110
2003	131,055	153,976	106,333
2004	135,891	161,012	108,864
2005	143,343	170,799	113,828
2006	152,289	179,385	123,022
2007	161,855	190,589	130,681
2008	160,422	188,616	129,650
2009	160,014	187,722	129,595
2010	159,607	186,592	129,812
2011	160,060	186,711	130,477
2012	160,364	186,609	131,100
2013	160,960	186,912	131,904
2014	161,391	187,123	132,512
2015	161,500	187,368	132,527
2016	161,387	187,355	132,337
2017	160,963	187,165	131,731
2018	161,249	187,333	132,074
2019	161,830	187,136	133,244
2020	161,779	186,391	133,743

Note: The results are for six education categories.

Table C.7 Real rural average human capital, 1985-2020 unit: yuan

year	real rural average human capital	male	female
1985	21,856	22,781	20,843
1986	22,018	23,411	20,490
1987	22,269	23,891	20,489
1988	22,517	24,726	20,079
1989	22,655	25,186	19,850
1990	22,921	26,027	19,475
1991	23,409	27,315	19,077
1992	24,160	28,501	19,346
1993	25,728	30,873	20,020
1994	27,499	33,028	21,370
1995	28,340	34,453	21,590
1996	30,256	36,887	22,924
1997	32,607	39,668	24,755
1998	34,199	41,799	25,705
1999	36,332	44,350	27,320
2000	38,896	47,442	29,228
2001	41,135	49,997	31,163
2002	43,461	52,399	33,438
2003	46,493	55,511	36,416
2004	50,040	59,077	39,968
2005	55,208	64,353	45,059
2006	59,796	67,846	50,821
2007	66,164	73,340	58,117
2008	66,125	73,086	58,257
2009	65,942	72,674	58,272
2010	65,829	72,348	58,349
2011	65,885	72,185	58,599
2012	65,770	71,919	58,611
2013	65,891	71,965	58,779
2014	65,958	72,042	58,816
2015	65,951	72,252	58,579
2016	66,083	72,592	58,485
2017	66,033	72,699	58,254
2018	66,513	73,133	58,729
2019	67,170	73,361	59,756
2020	67,255	72,656	60,603

Note: The results are for five education categories.

Table C.8 Real rural average human capital, 2000-2020 unit: yuan

year	real rural average human capital	male	female
2000	38,904	47,452	29,232
2001	41,145	50,010	31,169
2002	43,474	52,417	33,446
2003	46,512	55,533	36,429
2004	50,067	59,109	39,989
2005	55,248	64,398	45,094
2006	59,854	67,906	50,877
2007	66,248	73,421	58,205
2008	66,220	73,175	58,358
2009	66,047	72,772	58,385
2010	65,945	72,455	58,475
2011	66,011	72,301	58,738
2012	65,907	72,043	58,763
2013	66,039	72,098	58,945
2014	66,118	72,185	58,995
2015	66,121	72,404	58,770
2016	66,265	72,754	58,689
2017	66,226	72,871	58,471
2018	66,718	73,316	58,960
2019	67,388	73,554	60,005
2020	67,486	72,857	60,870

Note: The results are for six education categories.

Table C.9 Deflators used to adjust human capital

Year	CPI（1985=100）		Deflator for fixed capital formation (1952=1) (Zhang, 2004)	Deflators of fixed assets (2000=100) (Holz, 2006)
	Urban	Rural		
1985	100.00	100.00	1.28	34.6
1986	107.00	106.10	1.362	36.82
1987	116.39	112.70	1.434	38.75
1988	140.46	132.40	1.628	43.99
1989	163.34	157.90	1.766	47.73
1990	165.42	165.10	1.863	50.35
1991	173.85	168.90	2.021	55.13
1992	188.82	176.80	2.284	63.56
1993	219.23	201.00	2.856	80.47

Continue to the next page

1994	274.07	248.00	3.152	88.84
1995	320.12	291.40	3.34	94.08
1996	348.29	314.40	3.474	97.84
1997	359.09	322.30	3.533	99.51
1998	356.93	319.10	3.526	99.31
1999	352.31	314.30	3.512	98.91
2000	355.14	314.00	3.5506	100
2001	357.60	316.50	3.564802	100.4
2002	354.02	315.20	3.571932	100.6
2003	357.23	320.20	3.650515	102.81
2004	369.00	335.60	3.854943	
2005	374.89	343.00	3.916622	
2006	380.48	348.10	3.975372	
2007	397.62	366.90	4.130411	

Appendix D Calculation and selection of growth rate and discount rate

According to the income-based approach of calculating human capital index, human capital is computed from the discounted lifetime income.[24] In order to evaluate the lifetime income of a country, we need to estimate the lifetime income and adjust it by the survival rate. The future income of an individual with known gender and educational level is based on the average income of the group with the identical personal characteristics as this particular individual, and we must take into account the annual growth rate of real income.[25] We then convert the future income into current value according to the discount rate. Since we build the human capital indices respectively for urban and rural, we calculate the growth rate and discount rate separately.

1 Rate of Growth

1.1 Growth Rate of Real Income

The growth rates of real annual income are reported in the series of the *China Statistical Yearbook* published by National Bureau of Statistics of China. For urban areas, the average wage index divided by 100 is the growth rate of real wage. The wage only includes labor wage, which is defined as the average labor wage adjusted by inflation rate. 'Labor' refers to 'those who work in or get paid from the state-owned, urban collective, joint venture, joint-stock, foreign and Hong Kong, Macao and Taiwan invested in other units and its subsidiary bodies.' The average wage index

[24] Jorgenson, Dale W. and Barbara M. Fraumeni (1992b), "The Output of the Education Sector," in Z. Griliches, T. Breshnahan, M. Manser, and E. Berndt (eds.), The Output of the Service Sector, Chicago, NBER, 1992, pp. 303-341

[25] Jorgenson, D. W. and K – Y. Yun (1990). "Tax Reform and U.S. Economic Growth", *Journal of Political Economy* 98: pp.S151-193.

'reflects the relative change of real wage, indicating the degree that the level of real wage increases or decreases.[26] The calculation of average growth rate of real wage is given in Table D.1. Table D.1 shows the average growth rate is 7.09% over the period of 1978 – 2007; the trend of past thirty years is reported in Figure D.1.

For rural areas, we adopt, in general, the net income to evaluate the income status of farmers. According to the *China Statistical Yearbook*, 'net income' refers to 'the summed income from all sources deducted from the corresponding expenses'. The calculation formula is 'net income = total income – tax and fees – household operation expenses – depreciation of fixed assets used in productive activities – gifts to relatives.' The average net income of farmers is 'the level of net income with regards to the population, indicating the mean income of a region or a resident in a rural household.[27] After taking out the inflation effect, we find that the growth rate of net per cap income in rural areas is 6.34% from 1978 - 2007.

The above calculations, urban or rural, have obvious shortcomings: for urban areas, the wage is one of all possible sources of income, this method therefore is not comprehensive; for rural areas, the net income per cap includes all the family members in the household, thus it is not an accurate measure of the growth rate of productivity.

1.2 Rate of Labor Productivity

Harrod-Neutral model assumes the production function：

$Y=F(K, A(t)\cdot L(t))$

$A(t)=A_0 e(\theta t)$

$L(t)=L_0 e(nt)$

---

[26] All definitions here come from National Statistical Yearbook.

[27] National Bureau of Statistics of China, China Statistical Yearbook 2008, Website: http://www.stats.gov.cn/tjsj/ndsj/2008/indexch.htm.

Where A(t) is the measure of technological progress, A>0 and dA/dt>0. θ is the natural growth rate and n is the population growth rate. At the steady state, the growth rate of labor productivity (Y/L) and of real wage (w) equal to θ . Thus, Harrod-Neutral model provides the theoretical support the statement that we can use the growth rate of labor productivity to predict the future income.[28]

The real GDP is calculated as follows:

Real GDP = 1978 Nominal GDP * Real GDP Index (base = 1978)

Thus, the growth rate of labor productivity is calculated as follows:

Labor productivity = (1978 Nominal GDP * Real GDP Index)/ Employment

Growth rate of labor productivity for year *t* = ln (Labor productivity for year *t*) - ln (Labor productivity for year *t*-1)

According to the above method, the growth rate of labor productivity is 7.09% (Table D.1).

In order to calculate the rural and urban growth rate of labor productivity, we use primary industry for rural areas and secondary and tertiary industries for urban areas.

Labor productivity of agriculture is calculated as follows:

Labor productivity of Agriculture = Real GDP of Agriculture/ Employment of Agriculture

Real GDP is calculated in the same way. The rural and urban growth rate of labor productivity is 4.11% and 6.00% respectively.

Some research uses the growth rate of GDP per capita as the urban growth rate of labor productivity,[29] which is far greater than our estimation. Since they use population instead of employment, their method is less accurate.

---

[28] http://homepage.newschool.edu/het/essays/growth/neoclass/solowtech.htm

[29] Xu Xunchuan (2008), "The analysis of labor productivity's impact on employment." *Contemporary Finance & Economics* 10: pp.17-22

Following the above calculation, we draw the trend graph to show the change of the growth rate of labor productivity, the growth rate of real wage, and the rural and urban growth rate of labor productivity (Figure D.1 and D.2). As shown in Figure D.1, although the mean values of the growth rate of labor productivity and real wage are close, the growth rate of real wage varies dramatically. This fact indicates the statistical data is not stable across time. In Figure D.2, we notice that the rural growth rate is constantly lower than the urban growth rate. One possible reason is that service and industry have grown faster than agriculture during the past thirty years.

In conclusion, we choose 4.11% and 6.00% to estimate the lifetime income. After thirty years of economic transition, China's average growth rate is close to steady-state. In future research, we will apply time-varying growth rate to reflect the transformation of the economic structure.

1.3 International Comparison

According to the United States Bureau of Labor Statistics, the labor productivity estimates are 1.5% (U.S. 1979-2007), 2.0% (Japan), and 4.3% (North Korea). OECD use GDP per hour worked to measure labor productivity: 1.62% for U.S. from 1979 to 2007, 2.61% for Japan, and 5.29% for South Korea (Data for 1978 and 1979 are missing). The difference is due to the disparity of the length of working hours in each country. In addition, the labor productivity of Taiwan increased significantly: 7.38% during the period of 1953 – 1961, 9.15% during the period of 1962 – 1971, and 3.84% during the period of 1972 -1981.[30] The United States Bureau of Labor Statistics also published labor productivity

[30] Zhang Yushan(1987), "The comparison of labor productivity of Tiwan and South Korea." *Asia-pacific Economic Review* 6

in the non-agriculture sector: 1.4%-1.5% from 1979-1995 and 2.5% from 1995-2008.[31]

2 Rate of Discount

Discount rate reflects the time value of currency and is computed from the long-term rates of return. We use the rates of return of long-term government bonds as our discount rate. We choose the 10-year government bonds (1996-1997) and use its average interest rate. After taking out average inflation, the discount rate reported in Table D.3 is 3.14%. This is lower than the U.S. discount rate of 4.58%.[32]

---

[31] http://data.bls.gov/PDQ/servlet/SurveyOutputServlet and http://www.bls.gov/fls/#tables

[32] Jorgenson, D. W. and K – Y. Yun (1990). "Tax Reform and U.S. Economic Growth", *Journal of Political Economy,* 98: S151-193.

Tables and Figures of Appendix D

Table D.1 Growth Rate in China, 1978-2007

Year	Nominal GDP （100 million）	Real GDP Indices (1978 =100)	Real GDP （100 million）	Employed Person (10 thousand)	Labor Productivity (Yuan Per Person)	National Labor Productivity	Average Real Wage
1978	3645.22	100.00	3645.22	40152	907.85		
1979	4062.58	107.60	3922.25	41024	956.09	0.0518	0.0670
1980	4545.62	116.01	4228.75	42361	998.26	0.0432	0.0610
1981	4891.56	122.09	4450.47	43725	1017.83	0.0194	-0.0110
1982	5323.35	133.15	4853.54	45295	1071.54	0.0514	0.0150
1983	5962.65	147.60	5380.29	46436	1158.65	0.0782	0.0140
1984	7208.05	170.00	6196.81	48197	1285.72	0.1041	0.1470
1985	9016.04	192.89	7031.28	49873	1409.84	0.0922	0.0530
1986	10275.18	209.95	7653.29	51282	1492.39	0.0569	0.0830
1987	12058.62	234.27	8539.80	52783	1617.91	0.0808	0.0100
1988	15042.82	260.70	9503.13	54334	1749.02	0.0779	-0.0080
1989	16992.32	271.29	9889.27	55329	1787.36	0.0217	-0.0480
1990	18667.82	281.71	10268.92	64749	1585.96	-0.1195	0.0920
1991	21781.50	307.57	11211.50	65491	1711.91	0.0764	0.0400
1992	26923.48	351.37	12808.09	66152	1936.16	0.1231	0.0670
1993	35333.92	400.43	14596.65	66808	2184.87	0.1208	0.0710
1994	48197.86	452.81	16506.00	67455	2446.96	0.1133	0.0770
1995	60793.73	502.28	18309.27	68065	2689.97	0.0947	0.0380
1996	71176.59	552.55	20141.76	68950	2921.21	0.0825	0.0380
1997	78973.03	603.92	22014.35	69820	3153.01	0.0764	0.0110
1998	84402.28	651.23	23738.81	70637	3360.68	0.0638	0.0720
1999	89677.05	700.85	25547.66	71394	3578.40	0.0628	0.1310
2000	99214.55	759.95	27701.66	72085	3842.92	0.0713	0.1140
2001	109655.17	823.02	30000.98	73025	4108.32	0.0668	0.1520
2002	120332.69	897.77	32725.69	73740	4437.98	0.0772	0.1550
2003	135822.76	987.78	36006.57	74432	4837.51	0.0862	0.1200
2004	159878.34	1087.39	39637.85	75200	5270.99	0.0858	0.1050
2005	183217.40	1200.84	43773.17	75825	5772.92	0.0910	0.1280
2006	211923.50	1340.70	48871.43	76400	6396.78	0.1026	0.1271
2007	249529.90	1500.70	54703.78	76990	7105.31	0.1050	0.1360

Data Source：

1. Total employed person and average real wage growth rate of 1978-1990: 55 year's data of New China, Department of Comprehensive Statistics of National Bureau of Statistics of China, Beijing, China Statistics Press, 2005, P118-P119.
2. Other data: National Bureau of Statistics of China, China Statistical Yearbook 2008, Table 2-1, 2-2, 4-3, 4-23.
   Website: http://www.stats.gov.cn/tjsj/ndsj/2008/indexch.htm.

Note:

1. Indices of Gross Domestic Product (1978=100)：Real GDP index is the multiple of nominal GDP based on base GDP, which is calculated based on constant price. Here the base year indicates 1978.
2. Employed Persons refers to persons aged 16 and over who are engaged in gainful employment and thus receive remuneration payment or earn business income.
3. Average real wage growth rate equals to indices of average real wage growth rate (preceding year=100) divided by 100. Average real wage of staff and workers refers to the average wage of staff and workers after removing the effects of the price changes. Average real wage indices of staff and workers refers to the change of real wage, which reflects the relative increasing or decreasing level of real wage of staff and workers. Here wage only indicates wage of staff and workers; staff and workers refer to persons working in, and receive payment from units of state ownership, collective ownership, joint ownership, share holding ownership, foreign ownership, and ownership by entrepreneurs from Hong Kong, Macao, and Taiwan, and other types of ownership and their affiliated units.
4. Real GDP=Nominal GDP of 1978 * Indices of GDP(1978=100)
5. Labor Productivity Growth Rate= Ln (Labor Productivity of year *t*) - Ln (Labor Productivity of year *t*-1).

Table D.2 Growth rate of labor productivity of urban and rural sector

Year	Labor Productivity Growth Rate in Rural Sector				Labor Productivity Growth Rate in Urban Sector			
	Real GDP of Primary Industry (100 million)	Total Employed Persons of Primary Industry (10 thousand)	Labor Productivity of Primary Industry（Yuan Per person）	Labor Productivity Growth Rate of Primary Industry	Real GDP of Secondary and Tertiary Industry (100 Million)	Total Employed Persons of Secondary and Tertiary Industry (10 Thousand)	Labor Productivity of Secondary and Tertiary Industry (Yuan Per Person)	Labor Productivity Growth Rate of Secondary and Tertiary Industry
1978	1027.53	28318	362.86		2617.68	11835	2211.81	
1979	1090.21	28634	380.74	0.0481	2829.36	12391	2283.40	0.0319

Continue to the next page

1980	1074.39	29122	368.93	-0.0315	3141.99	13239	2373.29	0.0386
1981	1149.41	29777	386.01	0.0453	3285.95	13948	2355.86	-0.0074
1982	1281.93	30859	415.42	0.0734	3550.37	14436	2459.38	0.0430
1983	1388.66	31151	445.78	0.0706	3978.19	15285	2602.68	0.0566
1984	1567.53	30868	507.82	0.1303	4624.06	17329	2668.39	0.0249
1985	1596.43	31130	512.83	0.0098	5475.72	18743	2921.47	0.0906
1986	1649.41	31254	527.74	0.0287	6072.21	20027	3032.01	0.0371
1987	1727.00	31663	545.43	0.0330	6918.75	21121	3275.77	0.0773
1988	1770.94	32249	549.15	0.0068	7888.05	22085	3571.68	0.0865
1989	1825.40	33225	549.41	0.0005	8231.92	22105	3724.01	0.0418
1990	1959.16	38914	503.46	-0.0873	8467.07	25835	3277.37	-0.1278
1991	2006.18	39098	513.12	0.0190	9482.73	26393	3592.90	0.0919
1992	2100.49	38699	542.78	0.0562	11189.03	27453	4075.70	0.1261
1993	2199.24	37680	583.66	0.0726	13114.80	29128	4502.47	0.0996
1994	2287.22	36628	624.45	0.0675	15207.21	30827	4933.08	0.0913
1995	2401.60	35530	675.94	0.0792	17122.74	32535	5262.87	0.0647
1996	2524.11	34820	724.90	0.0699	19053.79	34130	5582.71	0.0590
1997	2612.44	34840	749.84	0.0338	21064.22	34979	6021.96	0.0757
1998	2703.85	35177	768.64	0.0248	22906.61	35460	6459.84	0.0702
1999	2779.56	35768	777.11	0.0110	24853.24	35626	6976.15	0.0769
2000	2846.27	36043	789.69	0.0161	27220.98	36042	7552.57	0.0794
2001	2925.97	36513	801.35	0.0147	29670.32	36512	8126.18	0.0732
2002	3010.82	36870	816.61	0.0189	32643.08	36870	8853.56	0.0857
2003	3086.09	36546	844.44	0.0335	36457.69	37886	9623.00	0.0833
2004	3280.52	35269	930.14	0.0967	40391.58	39931	10115.34	0.0499
2005	3452.11	33970	1016.22	0.0885	44969.25	41855	10744.06	0.0603
2006	3624.72	32561	1113.21	0.0912	50683.62	43839	11561.31	0.0733
2007	3758.72	31444	1195.37	0.0712	57342.36	45546	12589.99	0.0852

Data Source:

1. The data come from National Bureau of Statistics of China, China Statistical Yearbook 2008, Table 2-1、2-2、4-3.

   Website：http://www.stats.gov.cn/tjsj/ndsj/2008/indexch.htm.

Note:

1. Because of data accessibility and statistical accuracy, we use labor productivity of primary industry to measure labor productivity of rural sector, and use labor productivity of secondary and tertiary industry to measure labor productivity of urban sector, although there exist some primary industry in urban sector, secondary and tertiary industry in rural sector.

2. Primary industry refers to agriculture, forestry, animal husbandry and fishery and services in support of these industries. Secondary industry refers to mining and quarrying, manufacturing, production and supply of electricity, water and gas, and construction. Tertiary industry refers to all other economic activities not included in the primary or secondary industries.
3. Indices of Gross Domestic Product (1978=100)：Real GDP index is the multiple of nominal GDP based on base GDP, which is calculated based on constant price. Here, the base year indicates 1978.
   Real GDP=Nominal GDP of 1978 * Indices of GDP (1978=100)
4. Labor Productivity Growth Rate= Ln (Labor Productivity of year *t*) - Ln (Labor Productivity of year *t*-1).
5. In some years, the sums of employed person in three industries in table 2 are more than total employed person in table 1; in some other years, it is the opposite. The reason might be round off.
6. The article calculates the real GDP in the form of multiplication of real GDP indices and base GDP. The Statistical Bureau publishes the national and industrial real GDP indices (base year=1978) in the yearbook. It is possible that the summation of three industries' real GDP is unequal to the national real GDP due to the inconsistent GDP growth in different industries.

Table D.3 Discount rate

Year	Ten-year Bond Rate（%）	Average Ten-year Bond Rate (%)	Inflation Rate (%)	Discount Rate (%)
1996	11.83	11.83	8.31	3.52
1997	9.78	9.78	2.79	6.99
1998	5.50	5.50	-0.79	6.29
1999	3.33	3.33	-1.41	4.74
2000	2.87	2.87	0.42	2.45
2001	2.95 3.05	3.00	0.69	2.31
2002	2.54	2.54	-0.80	3.34
2003	3.02	3.02	1.20	1.82
2005	4.44	4.44	1.80	2.64
2006	2.80 2.92	2.86	1.51	1.35
2007	3.40 4.40	3.90	4.80	-0.90
	Average Discount Rate：			3.14

Data Source:

1. Individual-oriented ten-year treasury bond: China Financial Association, China Financial Yearbook 1997-2008, China Financial Yearbook Editorial, table 1.4.2

   China Securities Regulatory Commission:

   http://www.csrc.gov.cn/n575458/n4239016/n4239073/ n9321343/n9321457/9334474.html

   (data of 1999)

   http://www.csrc.gov.cn/n575458/n4239016/n4239073/ n8913123/n8913221/ 9332062.html

   (data of 2000)

   http://www.csrc.gov.cn/n575458/n4239016/n4239073/n8876669/n8876824/ 8881333.html

   (data of 2001)

2. Consumer Price Index:

   National Bureau of Statistics of China (NBS),China Statistical Yearbook 2008,table 8-2.

   Website:http://www.stats.gov.cn/tjsj/ndsj/2008/indexch.htm;

   Department of Comprehensive Statistics of NBS,"55-year Data Collection Since the Establishment of New China" Beijing,China Statistics Press,2005,page84 to page85.

Note:

1. Inflation rate = (year *t* price level – year *t*-1 price level)/year *t*-1price level.

   The price level here is the CPI based on 1978.

2. Discount rate=Individual-oriented ten-year treasury bond rate - inflation rate rate.
3. In 2001, the government issued ten-year account treasury bond targeted at all sorts of investors :

   ten-year treasury bond: 2.95%, twelve-year treasury bond: 3.05%.

4. In 2006, the government issued ten-year treasury bond targeted at all banks/exchange: (bonds can be purchased by individuals).

   three-year treasury bond: 2.80%, sixteen-year treasury bond：2.92%.

5. In 2007, the government issued ten-year treasury bond targeted at all banks/exchange: (bonds can be purchased by individuals).

   three-year treasury bond: 3.40%, ten-year treasury bond：4.40%.

Figures of Appendix D

Figure D.1 Growth rate of national labor productivity and average real wage

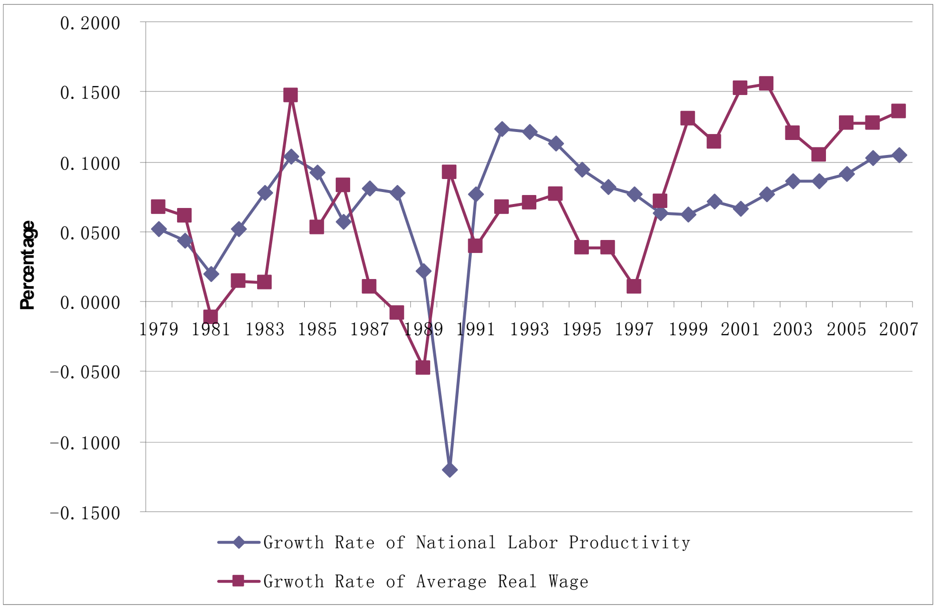
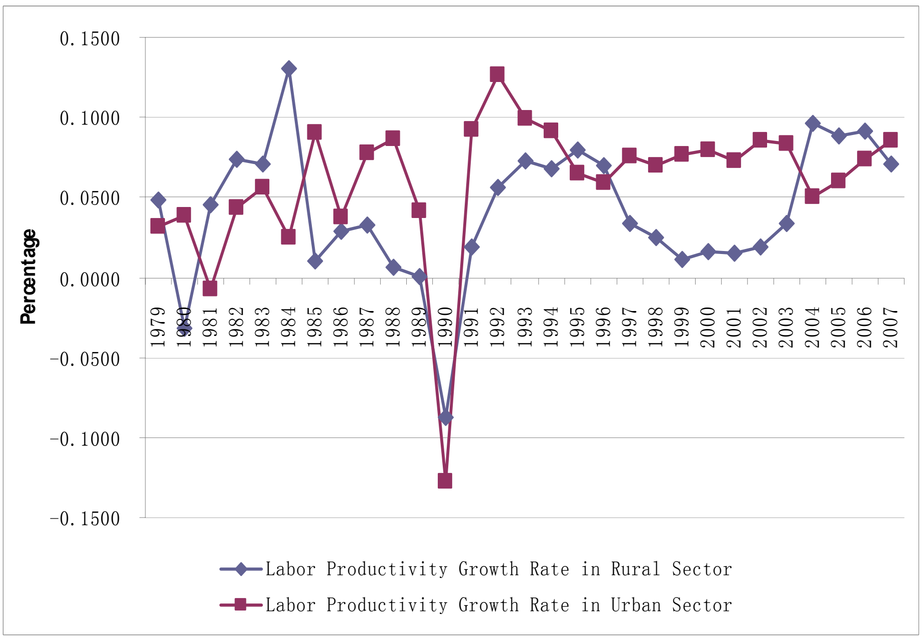


Figure D.2 Growth rate of labor productivity in urban and rural sectors



Appendix E Software package

In calculating China's human capital stock, we used a number of software including SAS, EXCEL, and STATA. Eventually, we developed a STATA platform that can be used to conduct all the computations. This STATA package comprises three parts: population estimation, population forecast, and human capital computation.

This package not only can be used to compute human capital at the national level, it can also be used at the provincial and local levels. The inputs required by the package are consistent with the data released by the National Bureau of Statistics of China. The package can also process supplemental data and parameters from microeconomic data sets or related research.

The package is flexible enough to adjust parameter values in order to assess the impact of different policy scenarios. We also provide a set of recommended values for all the parameters used in the calculation.

Acknowledgement

This research project and international symposium have benefited tremendously from supports inside and outside the Central University of Finance and Economics (CUFE). President Guangqian Wang of CUFE supported the project from the very beginning, and encouraged us to finish the preliminary study ahead of the schedule. Vice President Junsheng Li, President Assistant Lifen Zhao, and Director of Research Office Baowen Sun coordinated with various departments and facilitated the project and the symposium. Many offices in the university, such as the President's Office, Office of Research, Assets Management Office, Finance Office, Human Resources Office, and International Cooperation Office, cooperated to help the CHLR's research endeavor. Office spaces, conference facilities and computer labs were generously provided by the university Virtual Experiment Lab, the School of International Exchange, the Academy of Public Finance and Policy, and the School of Marxism Studies. The School of Economics at Georgia Institute of Technology, especially the Chair, Professor Patrick McCarthy, also generally supported the project. Last but not least, we greatly appreciate the help from two academic journals -- *Economic Research* and *Labor Economics and Relations*.