

**Educational Expansion and Wage Inequality: Evidence from urban China**

**by Wei Wang**

**(8081675)**

**Major Paper presented to the  
Department of Economics of the University of Ottawa  
in partial fulfillment of the requirements of the M.A. Degree  
Supervisor: Professor Mario Seccareccia**

**ECO6999**

**Ottawa, Ontario  
April 2017**

## **Abstract**

Despite the miracle of a dramatic increase in economic performance in the People's Republic of China (PRC), researchers are becoming increasingly concerned with income distribution and income inequality in China. Indeed, people have witnessed this growing inequality in the past two decades, and the tendency is towards a further increase. In this paper, a basic and standard human capital model is examined to understand this evolution of socio-demographic characteristics, which appear to be the reason for the widened gap of wage differential within China's labor market, especially after the educational expansion in 1999. Based on survey data from the Chinese Household Income Project (CHIP) conducted in 1995 and 2013, I decompose the difference in variance of the logarithm of wages into 1) endowment effect; 2) price effect; and 3) the unobserved effects and carry out micro-simulation technique to explain the evolution of wage inequality during that period. The results suggest that educational expansion targeting higher educational levels does not have an equalizing effect, although such a great change in the distribution of education can accelerate the pace of economic development. However, it could also generate unpredictable pressure on China's labor market and work in reverse to policy expectations.

## **Acknowledgement**

I would like to express my special thanks of gratitude to my supervisor Professor Mario Seccareccia for his contributions to all aspects of my work on this major research paper.

Secondly, I would also like to thank Professor Pierre Brochu for his helpful comments about this paper.

Finally, I need to thank my parents for their support and encouragement.

## Table of Contents

1. Introduction.....	5
2. Detailed Literature Review.....	9
3. Data.....	17
3.1 Data Description.....	17
3.2 Summary of socio-demographic characteristics.....	18
4. Empirical specifications and methodology.....	23
4.1 Model specifications.....	23
4.2 Decomposition.....	25
5. Empirical results.....	28
6. Conclusion and discussion.....	34
References.....	39
Appendix.....	41
Figure A.1 Distribution of Years of Schooling, 1995 and 2013.....	41
Figure A.2 Lorenz Curve for the Labor Force in China, 1995 and 2013.....	42
Table A.1 Log monthly wage equation, 1995 and 2013.....	43
Table A.2 Log monthly wage equation by age groups, 1995 and 2013.....	44
Table A.3 Proportion of different education level by age groups.....	45
Table A.4 Decomposition results.....	45

# 1. Introduction

Since the Chinese economy was reformed by Chairman Deng in 1978, the nation has gradually transformed itself from a planned economy to a more modern, market-based economy, which is still developing rapidly. At the time when the new policies were implemented, educational attainment did not play a very strong influence in the determination of wage earnings within the Chinese labor market. Rates of return to education had long been relatively low, concomitantly, for about 30 years prior, the educational profile of China's population likewise reflected this weakness. Mainstream neo-classical economic thinking could consider this an indicator of labor-market inefficiency. While that may well have been the case prior to the education reforms, nowadays, various economic transformations have generated far more appalling levels of income inequality in China which many Chinese analysts and commentators can hardly be indifferent. Indeed, the tremendous economic growth in China over the past two decades came with the continuing enlargement of income inequality and large changes to socio-economic demographics. In 2015, the National Bureau of Statistics (NBS) data reported that the Gini coefficient was 0.462 (National Bureau of Statistics of China (2016)).

Education, as a key element of human capital theory, has long been discussed as an important factor that can have a great impact on economic development and productivity, and also maybe a possible universal answer to most of our social problems (Mincer 1974). Certainly, the Chinese government paid much attention to educational improvements to change the domestic educational environment. In 1999, the Chinese government implemented a significant educational expansion policy which aimed to increase attendance of high school graduates to acquire post-secondary education after they finished their college entrance examination. As a result more people were educated, thereby contributing to a more sustainable growth in China's economic performance. According to the

China Statistical Yearbook from the National Bureau of Statistics (NBS), the numbers of entrants into formal post-secondary education (which includes regular undergraduate and college students) increased from 1.03 million in 1995 to almost 7 million in 2013. In 1999, right after this educational expansion policy went into effect, new entrants into tertiary education were half a million more than in 1998, a growth rate of 50 percent. Thereafter, the average increase in the volume of new entrants was approximately 390,000 per year. The range varied from a low of 73,000 (growth rate of 1.07 percent) between 2011 and 2012 and a high of 651,000 (42 percent) between 1999 and 2000. (National Bureau of Statistics of China, 2014)

A wide range of literature that explores the relationship between educational expansion and income distribution exists; the link between the two has been greatly debated. A critical view from Thurow (1975) questioned the implications of human capital theory and argued that education was not sufficient to explain the changes of earnings and earning inequality because the theory was based on a somewhat questionable view of the labor market that he described as the “wage competition” model. The wage competition hypothesis is based on the standard human capital theory that seeks to shed light on the developments of the domestic labor market. According to the standard wage competition hypothesis, the consequence of additional education achievement on individual income would suggest that a more equitable distribution of educational attainment would increase the supply of more educated workers, thereby reducing the supply of unskilled workers. This standard demand-supply mechanism would lead to a more equal income distribution as the wage gap between the more educated and the less educated would decline. An analysis based on demand and supply of skilled and unskilled labor markets, however, neglected to portray the full implication on the distribution of income. Despite the fact that education is more equally distributed, the differences in real wages between skilled and

unskilled workers are frequently enlarged and the rates of return to education do not necessarily have an equalizing or balancing effect. On the contrary, Thurow proposed a "job competition" model where education does not lead to the same impact on earnings as suggested by the wage competition model, with these earnings being mostly explained by on-the-job training and the consequences of education via its effect on the labor queue. According to Thurow, the labor queue is where education mostly takes effect. Potential workers are ranked within the labor queue, and the job competition/matching process is actually used to match trainable employees to employers. This potential effect of greater educational attainment can be to widen wage relativities.

Within this wide literature, however, many papers still rely on the demand and supply mechanism to explain the contribution of education to the structure of its labor market, which leads to further income inequality, and with most of these studies requiring an application of time-series data (Goldin and Katz 2007). Other studies based on standard human capital theory tend to regress the natural logarithms of earnings on a set of explanatory variables by different cohorts and to decompose (by using the 1973 Oaxaca-Blinder method) the change in inequality into composition and compression effects by groups or by different periods (Knight and Sabot 1983, Cortez 2001, Zhou 2014). Decomposition results usually vary among different countries, periods, and data used. According to Knight and Sabot (1983), composition and compression effects work in different directions, where the rates of return to education offset the growing composition effect and can result in a reduction in the overall income inequality. The worst-case scenario is where both impacts drive up the income inequality in the same direction (Zhou 2014).

Based on previous studies, this paper will examine and analyze the relationship between a large educational expansion in 1999 and wage inequality in urban China between 1995 and 2013. Two

micro-level cross-sectional data are drawn from Chinese Household Income Project (CHIP) 1995 and 2013 from the China Institute for Income Distribution. CHIP1995 is used as the base year representing educational attainments in the 1990s before the educational expansion, and CHIP2013 is the most up-to-date CHIP data set representing the current education profile, where monthly wages measured in CHIP2013 are adjusted to the Consumer Price Index from the National Bureau of Statistics. Moreover, CHIP2013 is able to include different waves of laborers who have benefited from educational expansion. Consequently, the education profiles and the returns in the two selected data should change significantly as compared to CHIP1995.

This article estimates the Mincerian logarithm wage earnings model and derives the Ordinary Least Squares (OLS) coefficient estimates for CHIP1995 and CHIP2013, with the interpretations of evidence to be found in section 5. Furthermore, the method of decomposition is used to separate the effects into 1) endowment effect; 2) price effect; and 3) unobserved effect and the decomposition method requires the construction of counterfactual simulations. Note that the sample utilized as part of this paper only contains observations pertaining to individuals aged from 25 to 64 who have wage earnings and are not retired, with observations on population outside of that age group are dropped.

The rest of this paper is organized as follows. In the following section, a literature review of existing studies on education and income inequality are discussed and summarized. Section 3 describes the data and summary of related variables generated to describe the change of evolution of China's labor market between 1995 and 2013. Thereafter, section 4 introduces the econometrics framework starting with the Mincerian human capital model, the necessary simulation methodology related to the topic and the decomposition. Section 5 then explains the results from the estimated models and the decomposition. Finally, Section 6 presents conclusions.

## 2. Detailed Literature Review

Many studies discuss income distribution and the link between income distribution or inequality and educational attainments in developed and less developed countries (LDCs). In this section, some of the previous papers is presented as an introduction to the theoretical framework of the topic, as well as empirical evidence in China and other countries, and econometric techniques that are useful in the field of income distribution and inequality.

As previously referred to, Thurow (1972, 1975) provides interesting explanations regarding education and income inequality using job competition theory and a critical view of the standard human capital theory based on the wage competition model of the labor market. The wage competition hypothesis illustrates the wages of the labor market of skilled workers and low-educated unskilled workers are determined by standard economic mechanism using demand and supply. Real wages yielded from a shift in either the demand-side or supply-side of the labor market reflects the wages of skilled and unskilled workers; therefore, the education profile of the labor force is able to alter the distribution of wages. With all the changes in educational attainments within society, education is expected to have an equalizing effect on wages or income distribution. He discusses the limitations of existing standard human capital theory and argues that education cannot fully explain the distribution of income, and an increase in education does not lead to a more equal distribution in the United States during the post-war period despite the more equal distribution of educational attainment. The author then introduces the job competition theory that assumes that the acquisition of on-the-job training gives rise to most job skills, and individuals' personal characteristics only serve as a screening device to match trainable employees to their entry-level jobs. That is, educational background does not directly act to alter the distribution of income, but it will change people's relative position or rankings in the

labor queue, which would probably enhance their chances of being matched at the port of entry of the training ladder (company). It is clear that the internal labor market, and the on-the-job training and experience acquired while on the job determines the wage distribution. With the job competition model of labor market, people may find themselves in an unfavorable position in the labor queue, when they are faced with a dramatic increase in the supply of educated workers. In this possible scenario, education becomes a necessity for them to maintain their current position in the labor queue. However it may not have any impact on earnings or the distribution of income as expected under the wage competition model. Finally, the author suggests that there is no direct evidence that greater education reduces the income inequality problem, and education itself may not be the possible universal solution to all of our socioeconomic problems or difficulties.

Battistón *et al.* (2014) examine the forceful impact of educational expansion on the measurement of income inequality for most Latin American countries. They used a micro dataset that combines the Socio-Economic Database for Latin America and Caribbean (SEDLAC) and World Bank's Latin America and the Caribbean Poverty and Gender Group for 18 Latin American countries from 1990 to 2009. They use a simulation technique to predict the change in the Gini coefficients if (1) education level in the initial year is simulated using the last year of educational attainment, and (2) the education level in the last year is replaced by its initial level of education. The results suggest that the education expansion has a non-equalizing effect. In particular, the simulated Gini coefficients for most Latin American countries show a positive change in Gini coefficients that vary from 0.2 to 2.5 point for simulation (1) and 0.7 to 3.2 points for simulation (2) over the selected period. Furthermore, the authors find that the impact of education expansion is still exhibiting a positive change in the Gini coefficient, but becoming smaller by using years of schooling rather than education levels.

Some researches focus on changes in the demand and supply mechanisms of the labor market structure, Goldin and Katz (2007) explore a different way of interpreting the changes in educational wage differentials by applying the supply, demand and institution (SDI) framework based on the time series data from 1915 to 2005, collected by Iowa State Census in 1915, 1940 to 2000 Census IPUMS, and 1980 to 2005 Current Population Survey (CPS) MORG samples in the U.S.. The framework is derived from the CES production functions with two factors: skilled, and unskilled laborers, where unskilled workers consist of high school graduates and high school dropouts. Relative wage premium between the skilled to unskilled as well as high school graduates to dropouts can be computed by log transformation of the CES production for the skilled to unskilled, as well as high school graduates to dropouts.<sup>1</sup> The relative wage is determined by relative demand, relative supply, and its elasticity of substitution between skilled and unskilled laborers. After estimating the SDI models, the results suggest that this model is performed well, and the simple demand and supply model would indicate that the growth in the relative supply of college (high school) workers to secondary school (high school dropouts) illustrate a distinct negative effect on the wage premium of the corresponding group. Goldin and Katz further decompose the contribution of the change in relative supply from immigrants and native-born Americans; and the effect of immigrants regarding the relative supply is much smaller than the native-born group.

Gregorio and Lee (2002) study the linkage between education and income inequality for 49 countries from 1960 to 1990 by using a cross-country unbalanced panel data. The estimated model of earnings at different education levels is a combination of earnings with no education and return to education (rate of return coefficient  $\times$  years of schooling) plus the error term, then taking the log

---

<sup>1</sup> For detailed definitions and equations see equation 1-4 of Goldin and Katz (2007)

transformation and the variance on both sides. The variable  $\text{var}(S)$  is a measure of educational inequality and it leads to more income inequality; on the other hand, education level  $S$  serves as an equalizer of earnings. The empirical evidence of their main findings supports the importance of education to income inequality. Higher education countries (10 OECD countries) tend to show less income inequality, that educational attainment of all of the four specifications have significant adverse effects on Gini coefficient, and educational inequality are positively correlated to income inequality as expected, as one standard deviation yield by decreasing in education inequality could effectively reduce the Gini coefficient by around 0.02 point. Authors also test the Kuznets inverted U hypothesis<sup>2</sup> by adding a log GDP per capita variable and its squared term in the human capital model. The result shows a positive effect of the log of GDP per capita and a significant negative effect of the square of the log of GDP per capita on income inequality, which means that income inequality rises as the economy grows and then eventually declines.

Chi *et al.* (2012) examine the natural logarithm of earnings, the year of entry into the labor market, and education expansion, which was driven by change in the supply of highly educated college graduates by using Chinese Urban Household survey from 1989 to 2009, the latter being a transition period from centrally planned to a more market-oriented economy. They analyze the performance of China's labor market by separating it into different cohorts for the first year at work, three, and five years of work experience by regressing the dependent variables (mean  $\ln$  earnings, standard deviation of  $\ln$  earnings,  $75P/25P$ , employment rate and percentage of white collar jobs) on the size of cohort and entry year log of GDP. Their results demonstrate that cohort size (a proxy for supply change) has a negative impact of -0.908 on the log of earnings, especially for college graduates starting their first

---

<sup>2</sup> See Kuznet (1955)

jobs, which was statistically significant at the 1 percent level. It eventually falls after on-the-job training is acquired. The standard deviation of ln earnings is used to measure the dispersion of earnings within the cohort. The cohort size contributes more to the wage differential for the first year at work, and becomes less in measured magnitude of coefficients as more experience is gained, with most of the results from students who have five years of work experience being statistically insignificant. Thus, supply change due to education expansion puts them into less favorable condition as the size of college graduates are huge compared to before.

Knight and Sabot (1983) use three samples from two low-income countries, Tanzania in 1971 and 1980 and Kenya in 1980, to evaluate the change in income inequality with respect to the increasing relative supply of educated individuals. They adopt the simulation technique to measure two effects that cause the change in income inequality, namely, composition and compression effect.<sup>3</sup> They simulate the earnings of the Kenya sample in 1980 by using rates of return to education (compression) and education structure (composition) in the 1980 Tanzanian sample. The results show that both effects are important in explaining income inequality. Knight and Sabot illustrate the changes of both effects and overall force in a plot with the measure of income inequality on the vertical axis and effects on the horizontal axis. The structural changes of education of the Tanzania 1971 are simulated using Kenya 1980 and they suggest an enormous increasing impact on income inequality from log variance of 0.23 to around 0.40. The effect of rates of returns to education tend to decrease and offset the increase in the composition effects resulting in a change of approximately -0.20 in log variance of income in the overall income inequality of total changing in educational coefficients and composition effect. They

---

<sup>3</sup> Both effects are measured by using the estimated coefficients and vector of educational (dummy) variables from the earning equation from different periods or different countries. Composition effect is the change in the distribution educational level while holding coefficient estimates (rate of return to education) unchanged. Similarly, compression effect is to use the new level of rate of return to education instead of the initial one and to predict the earnings with institutional structure (of a specific country or time) fixed at the original level.

argue that a shift from low-income to high-income urban areas would not lower the income inequality due to imperfections of the labor market and the accelerating accumulation process of human and as well as physical capital.

Following a similar procedure as Knight and Sabot (1983), Cortez (2001) studies the impact of relative supply change due to educational expansion on income inequality in Mexico, using datasets from the National Survey of Households' Income-Expenditure (ENIGH) for 1984, 1989, 1992, and 1996. Cortez also use a Mincerian log-linear specification to estimate the model and simulation technique for different time periods. The simulation results suggest that both effects are negative for female workers; compression effect for male workers rose from 0.299 in 1984 to 0.329 in 1996, whereas composition effect for male went to 0.235 in 1996 from 0.299. However, the overall effect for males decreased inequality from 1984 to 1996. The result is different from Knight and Sabot (1983), whose impact of the compression effect is outweighed by the composition in the case of Mexico. That is, educational expansion does not sufficiently explain the rise in income inequality (measured by the Gini coefficient, 0.431 in 1984 and 0.535 in 1996) in Mexico. Furthermore, Cortez uses the same approach to analyze the process of unionization, and it turned out to be explained by the relative change in income inequality. The simulation results of unionization for men increased from 0.3 to 0.349 and from 0.408 to 0.496 for women with an increase in both composition and compression effects.

Zhou (2014) analyzes the Chinese labor market structure and examines how the changes in the Chinese labor market contribute to the raising of earnings disparities in urban China by adopting Life History and Social Changes in Contemporary China (LHSCCC) in 1996 and Chinese General Social Survey (CGSS) in 2010. The model is primarily based on estimation of the log earnings on a set of education binary variables containing six levels of educational attainments and all others

corresponding socio-economic demographics including: female, age, squared age, party membership, rural hukou,<sup>4</sup> state sector (reference group), private sector, self-employment and provincial dummy variables combined in the dataset. The education dummy variables include no schooling, elementary school, junior high school, high school or vocational high school (reference group), vocational college, and four-year college (formal university or undergraduate). The methods that the author used in this paper also include: the variance of the log earnings regression, and a decomposition of composition and non-composition effects that contributed to the earnings inequality that requires a further counterfactual analysis.<sup>5</sup> Compositional changes have an effect of shifts in the distribution of the set of education variables from 1996 to 2010, while holding the price of returns to education constant, and an allocation effects of changes in distribution of socio-economic status from 1996 to 2010, with rates of return to all other socio-economic variables are assumed to be fixed. Non-compositional changes, a similar term for "compression effect" in Knight and Sabot (1983), and Cortez (2001) are the changes in the corresponding coefficient estimates of education dummies and socio-economic demographics from 1996 to 2010, holding the distribution of them unchanged. The result shows the earnings inequality measured by variance increased to about 0.304 from 1996 to 2010, and this rise in inequality is explained by the non-compositional change that largely contributed by price returns to education  $\beta_{edu}$  (45.2 percent), and between-group difference (45.8 percent), that is the variance of distribution of matrix X related coefficient estimates. For compositional effect, both distribution of education and the distribution of sector explained 41.9 percent of the entire difference in variance and total distributional effect of both education and socio-economic demographics are 54.9 percent of total

---

<sup>4</sup> Hukou status is the Chinese household registration system that aims to distinguish between permanent urban and rural residents. Zhou (2014) controls this binary variable because hukou discrimination may lead to earning differential among different hukou status

<sup>5</sup> Please see Zhou (2014) pp437-441 for detailed econometric model and counterfactual simulations used for analysis

variations in inequality. Zhou's analysis suggests that the rise in earnings inequality from 1996 to 2010 can be due primarily to both price effect and distribution of educational attainments, as the rates of return to education, and the change of educational structure in China's labor market are altered significantly. This was further compounded because of the change in the sectoral structure and movement from state-owned enterprise (SOE) or public sector to private sector after transition from a centrally planned economy to a more market-oriented structure.

The decomposition and microsimulation methodology of this paper are primarily based on Bourguignon *et al.* (2005) and Bouillon *et al.* (2003). They adopt different decomposition methodology which is primarily based on Oaxaca (1973) and Blinder (1973) decomposition method, although the Oaxaca-Blinder method is often used to isolate the difference in means between two cohorts or two selected periods, which could account for the discrimination of one group against the other group captured by the gap of different means in the matrix  $X$  and various means in its related coefficient estimates for the two groups but not the distribution between them. Bouillon *et al.* (2003) examine income inequality of Mexican households from 1984 to 1994 using National Household Income and Expenditure Surveys (ENIGH). Their decomposition results suggest that there is a great proportion of income inequality that can be explained by the price effects, which account for 73.18 percent of the total shifts in the Gini index. The methodology of decomposition and micro-simulation will be discussed in detail in section 4.

Given the existing large volume of previous literature on education and income inequalities, some researchers provide evidence on how educational improvement can further reduce income inequalities, while others have the opposite answers regarding increasing education level. In addition, some papers develop and provide decomposition technique based on the most adopted Oaxaca-Blinder method, but

different than the regular Oaxaca-Blinder that is restricted to decompose between the mean values of the two separate groups. Unfortunately, researchers and scholars have not reached one clear answer to the enlarged wage inequalities; evidence may vary because of various factors that differ across countries. In order to have a deeper understanding about educational expansion and wage inequalities, this paper examines descriptively and empirically the effects of education and wage inequalities for urban households before and after the implementation of educational expansion in China in 1999. The feature of such policy can possibly provide some useful information that contributes to the topic of wage inequalities in China, especially during the transition period. The method used in this paper is primarily based on the simulation techniques and decomposition method followed by Bourguignon *et al.* (2005) and Bouillon *et al.* (2003) to examine influence of education on wage differentials.

### **3. Data**

#### **3.1 Data Description**

In this paper, an extensive micro-level cross-sectional data is drawn from the Chinese Household Income Project (CHIP) 1995 from China Institute for Income Distribution to be used as the base level of performance in China's labor market; CHIP2013 is used to analyze the evolutionary changes of income inequality in China after an outstanding educational expansion implemented in 1999. The Chinese Household Income Project has a total of five surveys conducted in 1989, 1996, 2003, 2008 and 2013. CHIP 2013 is the latest survey that is available, and it allows the potential observations who benefited from the education expansion to enter the labor market 14 years after finishing their 3-year college or 4-year university programs.

The CHIP datasets have rich, detailed information for income inequality analysis including sociodemographic characteristics, employment status, retirement status, and income information at the

individual level. The CHIP datasets are unweighted raw survey data, thus the size of CHIP samples is not fully representative to the overall Chinese national population. In order to make it representative, this paper follows sample weights calculation in Song, Sicular and Yue (2011) by using data published by the National Bureau of Statistics (NBS) of urban population information from the Chinese 2000 and 2010 censuses.

For research purpose, the two samples used in this paper only consist of people located in the urban areas who are considered wage earners, not retired, full-time employment between the ages of between 25-64 at the time of both CHIP 1995 and CHIP2013 conducted in the early 1996 and 2013. The CHIP dataset also contains household information for each household members. There are a total of 10,455, and 9,384 observations that satisfied the restrictive conditions mentioned above corresponding to CHIP1995 and CHIP2013. For the selected sample in 1995 and 2013, samples used in this paper only account for 1) head; 2) spouse; and 3) child, and other household members are dropped in the descriptive results as their educational choices may be irrelevant to the household and the observations are not sufficient. Table 1 shows the proportion of household members in CHIP1995 and CHIP2013.

Table 1 Proportion of household members in the two samples

Household code	CHIP 1995			CHIP 2013		
	freq.	percent	cum.	freq.	percent	cum.
(1) head	5,132	50.03	50.03	4,354	49.13	49.13
(2) spouse	4,592	44.77	94.79	3,476	39.22	88.35
(3) child	534	5.21	100	1,032	11.65	100
Total	10,258	100		8,862	100	

Source: author's calculation

### 3.2 Summary of socio-demographic characteristics

Table 2 Evolution of characteristics of the labor force in China, 1995 and 2013

	Total			Men			Women		
	1995	2013	%Δ	1995	2013	%Δ	1995	2013	%Δ
<b>monthly wage</b>	495.06	1918.06	287.44	535.95	2143.78	300.00	448.59	1625.00	262.24
<b>log monthly wage</b>	6.06	7.31		6.15	7.44		5.95	7.14	
<b>Age Structure (in percentage)</b>									
			Δ			Δ			Δ
25-40	52.32	46.39	-5.93	47.46	41.78	-5.68	57.85	52.37	-5.48
41-54	42.54	45.31	2.77	44.16	46.60	2.44	40.70	43.63	2.93
55-64	5.14	8.30	3.16	8.38	11.62	3.24	1.45	4.00	2.55
<b>Education (in percentage)</b>									
			Δ			Δ			Δ
junior high	36.07	31.97	-4.1	33.05	32.05	-1.00	39.50	31.87	-7.63
high school	55.51	29.83	-25.7	55.80	30.00	-25.80	55.19	29.60	-25.59
college	8.42	38.20	29.8	11.15	37.95	26.80	5.31	38.53	33.22
<b>Average years of schooling by groups</b>									
			%Δ			%Δ			%Δ
age 25-40	11.09	12.90	16.33	11.46	12.92	12.74	10.76	12.89	19.84
age 41-54	10.35	10.86	4.91	10.65	11.08	3.96	9.98	10.57	5.86
age 55-64	9.43	9.18	-2.62	9.96	9.66	-3.00	9.02	7.32	-18.76
Head	10.93	11.56	5.78	11.07	11.40	3.00	10.64	12.11	13.78
Spouse	10.46	11.45	9.48	11.02	11.92	8.19	10.12	11.25	11.12
Child	11.47	12.73	11.07	11.41	12.39	8.54	11.58	13.34	15.22
<b>Wage ratio</b>									
			%Δ			%Δ			%Δ
90th / 50th	1.80	2.14	19.11	1.78	2.06	15.51	1.80	2.03	12.92
50th / 10th	2.03	2.39	17.37	1.88	2.16	15.04	2.12	2.50	18.16
90th / 10th	3.66	5.11	39.80	3.34	4.44	32.88	3.81	5.08	33.42
college/HS	1.23	1.37	11.39	1.20	1.37	14.51	1.23	1.37	11.57
HS/JS	1.15	1.31	13.88	1.08	1.23	14.39	1.22	1.44	18.54
college/JS	1.41	1.79	26.85	1.29	1.69	30.99	1.50	1.98	32.25

Source: author's calculation based on CHIP1995 and CHIP2013.

We start by looking at the characteristics of the labor force in China over the period 1995-2013 in Table 2, which could potentially affect the distribution of wages between 1995 and 2013. The most significant change is that the mean monthly wage has almost tripled (that is more than 10 percent per annum) in the past two decades between 1995 and 2013. The selected sample population is becoming older; the age structure of 25 to 40 decreased from 52 percent in 1995 to 46 percent in 2013 with a slight increase of the middle age group labor force, whereas the age of 55 to 64 increased from 5 percent to 8.3 percent.

Educational attainments are measured in both the highest level of education completed and years of schooling by the individuals in CHIP1995 and CHIP2013. The highest degree earned is then combined to three commonly used subgroups: junior high school, high school, and college or higher

degrees.<sup>6</sup> The distribution of education in term of highest degree earned by an individual is consistent as expected, with the junior high and high school (unskilled) decreasing from a total of 92 percent to 62 percent, leading to an equivalent proportion of the population who became more educated over the period 1995-2013. In general, the distribution of education after the educational expansion was well performed; however, it is surprising that only about 4 percent is contributed by the subgroup of junior high school completed or lower education levels, whereas high school graduates provides a significant proportional change of the total population that shifted toward college education. It is consistent with the goal of educational expansion, namely that senior high school graduates have a higher probability to go to college after taking their college entrance exams.

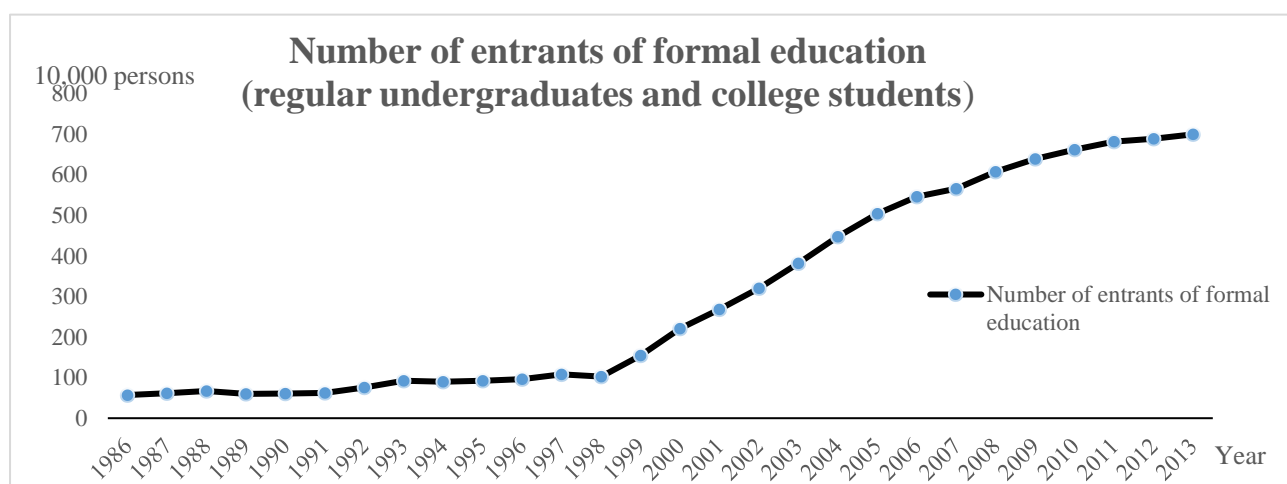
The educational expansion policy supposes to have a spillover effect to current secondary school students, although the descriptive result shows that it barely helped the bottom group, those who would no longer to get any further schoolings after having completed their 9-year compulsory education. There are various reasons for that, but unfortunately, that bottom group is not treated equally regarding the distribution of education. Furthermore, it is worth mentioning that more than 90 percent of the change in educational distribution at the bottom level (junior high school and elementary school) is due in large part to a decrease in the proportion of the female labor force at that level. It brings the concern about the relationship of unexpected limitation of this radical educational expansion (1999) aimed at high-level education and increase in wage differentials between the two selected periods. With the educational expansion opening more seats to the potential college and university students, it is impressive that the evolution of education distribution at the college or above category is increasing

---

<sup>6</sup> For detailed definitions of the three categories, junior high school includes individuals who were never schooled (including informal education such as literacy courses), elementary school and junior high school. High school consists of senior high school, vocational senior secondary school/technical school and specialized secondary school. College includes polytechnic college, undergraduate (Bachelor's degree) and graduate (Master's degree or above).

dramatically within the sample periods. The CHIP2013 dataset is enough for us to account for the observations of individuals who participated in the labor market and were benefitting from the education expansion implemented in the 1999 (see figure 1).

Figure 1. Education expansion measured by number of new entrants



Source: China Statistical Yearbook 2014 compiled by National Bureau of Statistics of China, of which the year 1996 from National Education Development report (1996) conducted by China Education and Research Network and year 1997-1999 based on data from Ministry of Education of the People's Republic of China.

It is also impressive if we evaluate education according to years of schooling. The average years of schooling by groups (age and household members) are also reported in Table 2. For the age groups, the younger age subgroup from 25 to 40 experienced a significant increase in average years of schooling from 11 in 1995 to around 13 in 2013; that is a growth rate of 16 percent.

Females between age subgroup 25 and 40 grew faster than the male labor force, with a total increase of more than two years on average compared to 1.5. The average years of schooling between men and women seems to become more equal only for the subgroup 25 to 40 who were highly likely to be influenced by the education expansion, whereas the other two age subgroups illustrate no change or slight change with respect to years of schooling. The distribution of years of education in 1995 and 2013 are plotted in the Figure A.1 in the Appendix. For CHIP1995, the figure has a smooth pattern; most of the workers in the labor force are junior high school graduates with 9 years of schooling, and

senior high school completed with 12 years of schooling. The two tails are narrowed, 4-year college graduates with 16 years of schooling is about 6 percent of the sample, and there are more junior or senior high school dropouts. On the other hand, CHIP2013 illustrates less dropouts and more college graduates, which could potentially cause a dramatic changes in the growing relative supply of highly educated employees.

The entire household group demonstrates positive growth for years of schooling on average. Heads of household, as demonstrated in the case of the female labor force, have a greater increasing level of educational attainments; female heads of household display a significant increase compared to the male head, with an average of one more year of schooling. The overall gap of average years of schooling between head and spouse is considerably small; the average spouse years of schooling advanced from 10.4 years in 1995 to 11.4 years in 2013. Although the percentage of children in both samples is relatively small compared to others, this sample only includes wage earners; therefore, a person who is of age 25 to 64 and is the child of the head could be considered a part of the labor force of the new or younger generation. On average, they tend to have higher educational attainment relative to their parents, and the educational attainment of females surpasses that of males in both samples of CHIP1995 and CHIP2013.

The wage ratios illustrate a simplified way to see the change of evolution of wage structure in percentiles from 1995 to 2013. We use 10<sup>th</sup>, 50<sup>th</sup> (median) and 90<sup>th</sup> percentile to represent the low wage, medium wage and high wage group, the increased wage ratio for the 90th percentile wage to 50th percentile and 10th percentile; and their percent changes during the past two decades support the enlargement of wage inequality. The wage ratio of 90/50 is 1.80 in 1995 and 2.14 in 2013 with a 19 percent increase. The 90/10 is 3.66 in 1995 and 5.11 in 2013 with an approximately 40 percent increase.

Alternatively, Lorenz curves of CHIP 1995 and CHIP2013 are presented to demonstrate visually the rise in inequality from 1995 to 2013; the widening gap between the 45-degree line and the other three curves suggests the enlargement of the wage gap between different percentile groups (see Lorenz curves of CHIP1995 and CHIP2013 in Appendix, Figure A.2). The wage premium of college graduates to high school and junior high school graduates increased by 11.39 percent and 26.85 from 1995 to 2013, and wage premium of high school to junior high school or below rose by 13.88 percent from 1995 to 2013. Even with a greater supply change of skilled workers, even with the wage premium and wage ratio still illustrate an enlargement of the wage gap between wage groups and education groups.

## 4. Empirical specifications and methodology

### 4.1 Model specifications

The model is based on the standard Mincerian logarithm earnings equation (Mincer, 1974),

$$\ln m\text{wage}_i = \alpha + \beta_1 \text{male}_i + \beta_2 \text{Han}_i + \beta_3 \text{party}_i + \beta_4 \text{exp}_i + \beta_5 \text{exp}_i^2 + \delta \text{junior}_i + \phi \text{college}_i + \mathbf{p}\boldsymbol{\gamma} + \varepsilon_i \quad \forall i = 1, \dots, n \quad (1)$$

$$\ln m\text{wage}_i = \alpha + \beta_1 \text{male}_i + \beta_2 \text{Han}_i + \beta_3 \text{party}_i + \beta_4 \text{exp}_i + \beta_5 \text{exp}_i^2 + \delta \text{schooling}_i + \mathbf{p}\boldsymbol{\gamma} + \varepsilon_i \quad \forall i = 1, \dots, n \quad (2)$$

where  $\ln m\text{wage}_i$  presents natural logarithm of the monthly wage for individual  $i$ . The model consists of three labor-market discrimination variables: (1) Male is a binary variable that equals to one if the individual  $i$  is male, and zero otherwise; (2) Han is a dummy variable that equals to one if individual  $i$  is Han, and zero if individual  $i$  is from ethnic minorities; (3) Party is a binary variable indicating individual's political status that equals to one if individual  $i$  is China Communist Party membership and zero if not. These are dummy variables that control for gender, ethnic groups and Chinese Communist Party membership (Appleton, Knight, Song, Xia 2009) for each individual  $i$ , namely those characteristics which could have an impact on an individual wage and possible

discrimination.  $exp_i$  and  $exp_i^2$  is the work experience and squared work experience for individual  $i$ . Note that experience in the CHIP1995 dataset is computed as age minus years of schooling subtracted by six, whereas CHIP2013 includes information on the year the individual  $i$  first entered the labor force; thus the work experience is measured as the individual's age deducted by his/her year of entry.

Education explanatory variables in equation (1) includes three common categories of educational attainments that were mentioned in the previous section and footnote: (1) junior is a binary variable that equals to one if the person stopped school after completion; (2) high school includes high school, vocational senior secondary school/technical school and specialized secondary school that are included in the constant term and it refers the reference group; (3) college is a binary variable that equals to one if the person attended a community college or higher level (undergraduate, graduate) and zero otherwise. The coefficient estimates  $\delta$  and  $\phi$  capture the average rates of return to junior high school and college shift from the benchmark group vis-à-vis the others. Schooling represents years of schooling, it varies from no education to 24 years of education in CHIP1995 and CHIP2013, and however, the distribution of years of schooling narrows at both tails (see Figure A.1). Correspondingly, the parameter  $\delta$  in equation (2) estimates the rate of return of an additional year of schooling.

We also controlled the provincial dummy variables to account for the wage differentials at the provincial level, as regional inequality tends to be significant in both rural and urban areas in China (Lu *et al.* 2010).  $\mathbf{p}_{1995}$  is a vector of 10 provincial dummy variables selected in CHIP1995, which includes Beijing Municipality, Shanxi, Liaoning, Jiangsu, Anhui, Henan, Hubei, Guangdong, Sichuan, Yunnan, Gansu. Similarly,  $\mathbf{p}_{2013}$  includes the same 10 provincial dummy variables in CHIP1995 with three additional provincial binary variables consisting of Shandong, Hunan, and Chongqing Municipality in CHIP2013. Beijing Municipality, the capital city of China is used as the reference term.

## 4.2 Decomposition

The decomposition method follows Bourguignon, Ferreira, and Lustig (2005) and Bouillon *et al.* (2003) extended from the Oaxaca-Blinder method (Oaxaca 1973, and Blinder 1973). Consider the Mincerian log-linear earnings equation  $\log \mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$  similar to equation (2) in the previous subsection 4.1. Distribution of wage  $\mathbf{y}$  in the year 1995 ( $t$ ) and 2013 ( $\tau$ ) is a function of  $\mathbf{X}$  including all explanatory variables in equation (2), rates of return related to independent variables  $\boldsymbol{\beta}$ , and unobserved factors included in the error term  $\boldsymbol{\varepsilon}$  for year= $\{1995,2013\}$ .

$$D(\mathbf{y}_t) = f(\mathbf{X}_t, \boldsymbol{\beta}_t, \boldsymbol{\varepsilon}_t) = (y_{1t}, \dots, y_{Nt}) = \{y_i\}^t \quad t=1995 \quad (3)$$

$$D(\mathbf{y}_\tau) = f(\mathbf{X}_\tau, \boldsymbol{\beta}_\tau, \boldsymbol{\varepsilon}_\tau) = (y_{1\tau}, \dots, y_{N\tau}) = \{y_i\}^\tau \quad \tau=2013 \quad (4)$$

The change in the distribution of wages is thus the difference between 1995 and 2013

$$\begin{aligned} \Delta D &= D(\mathbf{y}_\tau) - D(\mathbf{y}_t) = f(\mathbf{X}_\tau, \boldsymbol{\beta}_\tau, \boldsymbol{\varepsilon}_\tau) - f(\mathbf{X}_t, \boldsymbol{\beta}_t, \boldsymbol{\varepsilon}_t) \\ &= f(\mathbf{X}_{2013}, \boldsymbol{\beta}_{2013}, \boldsymbol{\varepsilon}_{2013}) - f(\mathbf{X}_{1995}, \boldsymbol{\beta}_{1995}, \boldsymbol{\varepsilon}_{1995}) \end{aligned} \quad (5)$$

We then separate it into three parts

$$\begin{aligned} &= [f(\mathbf{X}_\tau, \boldsymbol{\beta}_\tau, \boldsymbol{\varepsilon}_\tau) - f(\mathbf{X}_t, \boldsymbol{\beta}_\tau, \boldsymbol{\varepsilon}_\tau)] + [f(\mathbf{X}_t, \boldsymbol{\beta}_\tau, \boldsymbol{\varepsilon}_\tau) - f(\mathbf{X}_t, \boldsymbol{\beta}_t, \boldsymbol{\varepsilon}_\tau)] + [f(\mathbf{X}_t, \boldsymbol{\beta}_t, \boldsymbol{\varepsilon}_\tau) - f(\mathbf{X}_t, \boldsymbol{\beta}_t, \boldsymbol{\varepsilon}_t)] \quad (6) \\ &= [f(\mathbf{X}_{2013}, \boldsymbol{\beta}_{2013}, \boldsymbol{\varepsilon}_{2013}) - f(\mathbf{X}_{1995}, \boldsymbol{\beta}_{2013}, \boldsymbol{\varepsilon}_{2013})] \\ &\quad + [f(\mathbf{X}_{1995}, \boldsymbol{\beta}_{2013}, \boldsymbol{\varepsilon}_{2013}) - f(\mathbf{X}_{1995}, \boldsymbol{\beta}_{1995}, \boldsymbol{\varepsilon}_{2013})] \\ &\quad + [f(\mathbf{X}_{1995}, \boldsymbol{\beta}_{1995}, \boldsymbol{\varepsilon}_{2013}) - f(\mathbf{X}_{1995}, \boldsymbol{\beta}_{1995}, \boldsymbol{\varepsilon}_{1995})] \end{aligned}$$

$\Delta D = \log Y_{2013}$ total effect @2013	– price & unobserved effects of $\log Y_{1995}$	Endowment effect
+ p & u effects of $\log Y_{1995}$ @2013	– unobserved effects of $\log Y_{1995}$	Price effect
+ unobserved effects of $\log Y_{1995}$ @2013	– $\log Y_{1995}$ no effect	Unobserved effect

The first term in  $[\bullet]$  is the difference of income distribution of  $\mathbf{y}_{2013}$  and  $\mathbf{y}_{1995}$  at the level of  $\mathbf{X}=1995$ .

As we expect, education expansion should have a positive effect on average years of schooling, this term captures the endowment effect, with  $\mathbf{y}_{2013}$  at  $\mathbf{X}=1995$  should be less than  $\mathbf{y}_{2013}$  while other

factors remain unchanged. Then similarly, we take  $y_{2013}$  at  $X=1995$  minus  $y_{2013}$  at  $X=1995$ , and vector  $b=1995$  (coefficient estimates) is the price (rates of return) effect. The last term is the unobserved effect,  $y_{2013}$  at  $X=1995$  and vector  $b=1995$  and  $\epsilon=2013$  (or equivalently  $y_{1995}$  is simulated when  $\epsilon=2013$ ) minus  $y_{1995}$ .

Moreover,  $f(X_t, \beta_t, \epsilon_t)$  and  $f(X_t, \beta_t, \epsilon_t)$  (the highlighted factor of equation (6)) are hypothetical which are not directly observable, and then they can be computed by using counterfactual simulation using the following equations:

$$f(X_t, \beta_t, \epsilon_t) = \{y_i\}^{2013} |_{X=X_{1995}} = \ln \widehat{y_{2013}} |_{X_{1995}} = \widehat{\alpha}_{2013} + X_{1995} \widehat{b}_{2013} + \widehat{\epsilon}_{2013} \quad (7)$$

$$f(X_t, \beta_t, \epsilon_t) = \{y_i\}^{2013} |_{X=X_{1995}, b=1995} = \ln \widehat{y_{2013}} |_{X_{1995}, b_{1995}} = \widehat{\alpha}_{1995} + X_{1995} \widehat{b}_{1995} + \widehat{\epsilon}_{2013} \quad (8)$$

For the endowment effect of  $X$ , we can continue to isolate the effects that were contributed by educational attainment measured in term of years of schooling, experience, and remainders (this should consist of all other explanatory variables except education and experience related variables) as in the followings:

$$\{y_i\}^{2013} |_{s=s_{1995}} = \ln \widehat{y_{2013}} |_{s_{1995}} = \widehat{\alpha}_{2013} + X_{2013} \widehat{b}_{2013} + (s_{1995} - s_{2013}) \widehat{\delta}_{2013} + \widehat{\epsilon}_{2013} \quad (9)$$

$$\{y_i\}^{2013} |_{exp=exp_{1995}} = \ln \widehat{y_{2013}} |_{exp_{1995}} = \widehat{\alpha}_{2013} + X_{2013} \widehat{b}_{2013} + (exp_{1995} - exp_{2013}) \widehat{b}_{1_{2013}} + \widehat{\epsilon}_{2013} \quad (10)$$

For simplicity, assume that  $Z$  matrix have all other (remaining) explanatory variables with a vector of coefficient estimates  $\tilde{b}$  for those variables other than education and experience that are still included in a  $n \times 2$  matrix  $\tilde{X}$  with a  $2 \times 1$  vector of coefficient estimates  $\widehat{\Psi} = \left\{ \begin{matrix} \widehat{b}_1 \\ \widehat{\delta} \end{matrix} \right\}$ .

$$\{y_i\}^{2013} |_{z=z_{1995}} = \ln \widehat{y_{2013}} |_{z_{1995}} = \widehat{\alpha}_{2013} + \tilde{X}_{2013} \widehat{\Psi}_{2013} + Z_{1995} \widehat{b}_{2013} + \widehat{\epsilon}_{2013} \quad (11)$$

Similarly, these can also be calculated for price effect.

$$\{y_i\}^{2013}|_{\text{exp}=\text{exp1995}} = \ln \widehat{\mathbf{y}}_{2013} | \text{exp}_{1995} = \widehat{\boldsymbol{\alpha}}_{2013} + \mathbf{X}_{2013} \widehat{\boldsymbol{b}}_{2013} + \text{exp}(\widehat{b1}_{1995} - \widehat{b1}_{2013}) + \widehat{\boldsymbol{\varepsilon}}_{2013} \quad (12)$$

$$\{y_i\}^{2013}|_{\text{exp}=\text{exp1995}} = \ln \widehat{\mathbf{y}}_{2013} | \mathbf{s}_{1995} = \widehat{\boldsymbol{\alpha}}_{2013} + \mathbf{X}_{2013} \widehat{\boldsymbol{b}}_{2013} + \mathbf{s}(\widehat{\delta}_{1995} - \widehat{\delta}_{2013}) + \widehat{\boldsymbol{\varepsilon}}_{2013} \quad (13)$$

In order to compute the required counterfactual simulation of the changes in educational attainment, we could not assign a fixed proportional increase to the years of schooling, because each observation should be treated differently based on his/her personal characteristics such as gender and province in which he/she is located.

$$S_{j,\text{year } t}^{\text{distribution of year 2013}} = (S_{j,t} - \bar{S}_{j,t}) \cdot \frac{\sigma_{j2013}}{\sigma_{j1995}} + \bar{S}_{j2013} \quad (14)$$

where  $\bar{S}$  and  $\sigma$  are the averages and standard deviations in the cluster  $j$ . For example, for the years of education group of 0-6 in 1995,

$$S_{0-6,2013}^{\text{distribution of year 1995}} = (S_{j,2013} - 4.888073) \cdot \frac{1.3054}{1.541206} + 5.121166$$

and we repeat this step for other groups using mean and standard deviation shown in the table below:

Education group	mean 1995	$\sigma$ 1995	mean 2013	$\sigma$ 2013
0-6	5.121166	1.305446	4.888073	1.541206
7--9	8.432909	0.7360574	8.529259	0.6209533
10--12	11.19094	0.8441656	11.51272	0.6782062
13-16	14.23992	1.021008	14.99125	0.949335
17 and above	17.79283	1.463184	18.15705	0.9577984

Moreover, distribution of the location of different groups of CHIP2013 will be reweighted according to the relevant data of CHIP1995. For example, for the distribution of log wage in 2013 and distribution of log wage in 1995 where  $\mathbf{X}=1995$ , the provincial dummy variable (included in  $\mathbf{X}1995$ ) will be reweighted according to the proportion of observations in the CHIP1995.

Note that  $\Delta D = D(\mathbf{y}_\tau) - D(\mathbf{y}_t)$  only captures the difference of distribution of wage between 1995 and 2013. Instead, we will calculate  $\text{var}[D(\mathbf{y}_\tau)] - \text{var}[D(\mathbf{y}_t)]$ , where the function  $\text{var}[\bullet]$  is the

inequality index measured in variance of the logarithm of wages computed at each step of the simulation equation.

## 5. Empirical results

In this section, Table 3 presents the estimated results of the wage earnings equation for CHIP1995 and CHIP2013 and Table 4 interprets the coefficient estimates outcome of the wage earnings equation (1) and (2) by age group. Table 5 presents the decomposition results from the endowment, price and unobserved factors regarding changes in wage inequality.

Table 3 Log monthly wage equation, 1995 and 2013

Regression	1995 (1)	1995 (2)	2013 (3)	2013 (4)
Junior	-0.199*** (0.0118)	-	-0.214*** (0.0234)	-
High School	-	-	-	-
College	0.202*** (0.0146)	-	0.356*** (0.0213)	-
Years of Schooling	-	0.0418*** (0.0018)	-	0.0830*** (0.0034)
Experience	0.0373*** (0.0029)	0.0379*** (0.0029)	0.0342*** (0.0032)	0.0334*** (0.0031)
Experience squared	-0.000461*** (0.0001)	-0.000475*** (0.0001)	-0.0007523*** (0.0000)	-0.0006454*** (0.0000)
Han	0.0607*** (0.0231)	0.0592*** (0.0231)	0.0383 (0.0379)	0.0199 (0.0406)
Male	0.114*** (0.0101)	0.114*** (0.0101)	0.302*** (0.0169)	0.288*** (0.0169)
Party	0.084*** (0.0107)	0.0793*** (0.0108)	0.0485 (0.0209)	0.0534 (0.0205)
Constant	5.656*** (0.0419)	5.126*** (0.0481)	6.954*** (0.0457)	6.112*** (0.0518)
variance of residuals	0.2584	0.2642	0.4823	0.4766
adj. R-sq	0.279	0.2829	0.255	0.259
Number of observations	10,401	10,273	9,261	9,261

Note: dependent variable is natural log monthly wage, column (1) and (3) are equation (1) and column (2) and (4) are equation (2) for 1995 and 2013. Robust standard errors are in parentheses.

Provincial dummies are included but not reported, see Appendix for detailed information.

\* p<0.1, \*\* p<0.05, \*\*\* p<0.01

Source: author's calculation based on wage equation (1) and (2) from CHIP1998 and CHIP2013.

Table 3 presents the ordinary least square (OLS) estimates from equations (1) and (2) for CHIP1995 and CHIP2013. The results of Table 3 show a substantial change of rates of return to educational variables (dummy variable and years of schooling). In columns (1) and (3), the rate of return for having a junior high school or elementary school degree is -18.04 percent ( $\exp(-0.199)-1$ ) in 1995 and -19.2 percent in 2013 lower than the reference group (high school). More surprisingly, the rate of return for individuals having completed college education (and higher) yields 22.38 percent in 1995 and 42.76 percent higher relative to the benchmark category. In terms of years of schooling, the marginal effect for each extra year of schooling is reflected in a 4.18 percent increase in log wages for 1995 and a doubling of the coefficient (8.30 percent growth) in 2013. All of the rates of returns to education are statistically significant at 1 percent level.

Work experience and its quadratic term squared work experience have the expected signs; the marginal rate of return of experience for 1995 and 2013 are similar and stable, and they will increase at a diminishing rate. The majority of Han ethnic stands for around 95 percent of the sample in both CHIP1995 and CHIP2013; it presents 6.25 percent higher wages in advantage as compared to the minorities in our sample, and it is significant at the 1% level. However, this form of wage discrimination is no long statistically significant in CHIP2013. Similarly, the China Communist Party dummy variable is statistically significant in CHIP1995, but not in CHIP2013, although the effect of holding the status of Chinese Communist Party member, or the wage premium of CCP members are lower in 2013 than in 1995. After transitioning from the planned economy to a market-oriented economy, here we only control for the effects of CCP membership; however, the selection of party membership is also dependent on sociodemographic characteristics, and further analysis is beyond the scope of this paper (for further discussion see Appleton, Knight, Song, Xia (2009)).

The constant term refers to the reference group of people who have educational attainments at high school and its equivalent levels, who are in minority group, female and not Chinese Communist Party member increases significantly from 1995 to 2013 measured in the logarithmic term. In real terms, the monthly wage for the reference group is about 358.59 with a standard deviation of 190.8765 in 1995, and 885.60 with a standard deviation of 314.96 in 2013. In the complete regression result (Table A.1 in the Appendix), the table reports the provincial dummy variable, with most being statistically significant in relation to the reference group, Beijing, and only some of more highly developed provinces listed in the vector of provincial dummy variables performed relatively better than others (such as Jiangsu and Guangdong).

Table 4 Log monthly wage equation by age groups, 1995 and 2013

Regression	CHIP1995			CHIP2013		
	<35 (1)	35-50 (2)	>50 (3)	<35 (4)	35-50 (5)	>50 (6)
Junior	-0.237*** (0.0323)	-0.185*** (0.0129)	-0.175*** (0.0389)	-0.295*** (0.0647)	-0.188*** (0.0301)	-0.148*** (0.0398)
High School	-	-	-	-	-	-
College	0.286*** (0.0308)	0.155*** (0.0176)	0.208*** (0.0343)	0.278*** (0.0434)	0.374*** (0.0277)	0.456*** (0.0453)
Years of Schooling	0.0492*** (0.0045)	0.0372*** (0.0020)	0.0394*** (0.0048)	0.0683*** (0.0054)	0.0714*** (0.0034)	0.0811*** (0.0049)
Experience	0.0547*** (0.0102)	0.0503*** (0.0080)	0.0809*** (0.0125)	0.0547*** (0.0089)	0.0518*** (0.0038)	0.0608*** (0.0063)
Experience squared	-0.00138*** (0.0004)	-0.000708*** (0.0002)	-0.000957** (0.0002)	-0.00241*** (0.0000)	-0.000687** (0.0000)	-0.00113*** (0.0005)
Han	0.0581 (0.0415)	0.0844** (0.0298)	0.0435 (0.0654)	0.0453 (0.0759)	0.0671 (0.0529)	-0.1000 (0.1130)
Male	0.118*** (0.0204)	0.101*** (0.0120)	0.161*** (0.0455)	0.184*** (0.0318)	0.349*** (0.0216)	0.336*** (0.0515)
Party	0.0581** (0.0262)	0.0933*** (0.0119)	0.119*** (0.0310)	0.0182 (0.0364)	0.0259 (0.0269)	0.114*** (0.0467)
Constant	5.746*** (0.0870)	5.477*** (0.0993)	4.593*** (0.1947)	7.258*** (0.1221)	7.161*** (0.1767)	6.954*** (0.1792)
Variance of residual	0.3228	0.3184	0.4189	0.4134	0.3736	0.4625
adj. R-sq	0.2183	0.2470	0.3566	0.2113	0.253	0.335
Number of observation	2,878	6,309	1,214	2,386	5,308	1,567

Note: dependent variable is natural log monthly wage, wage earning equation (1) and (2) are regressed for different age groups. Robust standard errors are in parentheses.

The highlighted area of "years of schooling" is estimated by eqn. (2), otherwise it will be redundant  
Provincial dummies are included but not reported; see Appendix for detailed information.

\*  $p < 0.1$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

Source: author's calculation based on wage equation (1) and (2) from CHIP1998 and CHIP2013.

In this table, observations are divided into three age groups: new entrants or younger generation for individuals who are under 35 years old, middle age for age between 35 and 50, senior and experienced workers who are 50 and older. We re-estimated the wage earning equations (1) and (2) separately for observations belonging to the different age groups.

Interpretations of coefficient estimates are not different from those of the previous table, but there are some points worth mentioning about this table. First, the rate of return to education for individuals who have earned their college or higher degrees in 2013 is 32.05 percent, which is lower than its level 33.11 percent in 1995. It seems that educational expansion has generated enormous pressure on the labor market for new entrants or inexperienced laborers, that educational expansion influenced new entrants and may have actually exceeded labor market capacity, thereby bidding down their earnings. In the Appendix Table A.4, the proportion of education level over total population of both CHIP datasets by age shows its relative change in the structure of educational attainment; this could explain the stagnated rate of return to education for entry level new staff. For college completed wage earners under age of 35, the market share increased from 9 percent to 56 percent, and 6 percent to 34 percent of age between 35 and 50 years old. Second, the pattern of years of schooling between different age groups in 1995 reflects a stable distribution, although it has an increasing trend with respect to age even after controlling for experience. This also happens to the coefficient estimates measured for the dummy variable, it would suggest an effectively performed seniority provision within the internal labor market of the firm, where it could turn the new workers into the unfavorable minority (Thurow 1975).

Table 5 Decomposition results

Decomposition	Variance (LnY1995)	Variance (LnY2013)	Difference	%of total
Total effect	0.35237936	0.788367	0.4359876	100%
	$f(\mathbf{X}_{1995}, \boldsymbol{\beta}_{2013}, \boldsymbol{\varepsilon}_{2013})$	$f(\mathbf{X}_{2013}, \boldsymbol{\beta}_{2013}, \boldsymbol{\varepsilon}_{2013})$	Difference	% of total
Endowment	0.591388	0.788387	0.196999	45.18%
	$f(\mathbf{X}_{1995}, \boldsymbol{\beta}_{1995}, \boldsymbol{\varepsilon}_{2013})$	$f(\mathbf{X}_{1995}, \boldsymbol{\beta}_{2013}, \boldsymbol{\varepsilon}_{2013})$	Difference	% of total
Price	0.405722	0.591388	0.185666	42.59%
	$f(\mathbf{X}_{1995}, \boldsymbol{\beta}_{1995}, \boldsymbol{\varepsilon}_{1995})$	$f(\mathbf{X}_{1995}, \boldsymbol{\beta}_{1995}, \boldsymbol{\varepsilon}_{2013})$	Difference	% of total
Unobserved	0.35237936	0.405722	0.0533426	12.23%

Source: author's calculation based on CHIP1995 and CHIP2013, detailed calculations are available in the Appendix.

In Table 5 decomposition results, I have included the variance of log monthly wage of 0.3523 in 1995 and of 0.7884 in 2013, and its difference. Thus, following the decomposition method and required simulation process using equations (7) – (13), I decompose the resulting total change of inequality measure with respect to three effects. The endowment effect captures the modification of the distributional effect of a set of variables in the  $\mathbf{X}$  while holding coefficient estimates and unobserved factors unchanged. The price effect is the part that contributes to the change in the coefficient estimates of each variable and the constant term, where we fix the distribution of related variables and unobserved factors at the initial level. Similarly, an unobserved effect will be simulated holding all others constant. The table is based on equation (6), endowment effect accounts for the 45.18 percent of change in the total variations, and it is computed by using the distribution of wage earnings in 2013 (which is equivalent to the distribution of wage earnings in 1995 with  $\mathbf{X}$ ,  $\boldsymbol{\beta}$ ,  $\boldsymbol{\varepsilon}$  at the 2013 level, that is the joint effect of adding endowment, price and unobservable effects) minus the price and unobserved effects of the log monthly wage in 1995 at 2013 level. The difference gives the effect due to a change in distribution of the set of explanatory variables in  $\mathbf{X}$ . I proceed to break down the effect into detailed factors: education stands for a large proportion of endowment effect for about a half, whereas the other half relate to experience and remaining distributions of the labor market structure. By fixing the

distribution of X and price effect of CHIP1995, I obtained the price effect that yield from the difference between the joint price and unobserved effects subtract unobserved effect solely. The latter part contributes 42.59 percent of the change in total wage inequality, of which 42.49 percent of price effect is due to the change in returns to education over the selected period. Finally, the unobserved effect reports 12.23 percent of the rise in wage inequality that could be interpreted by unobserved individual skills or abilities that are not included in our model.

These counterfactual simulation and decomposition results suggest that the increment of wage inequality from 1995 to 2013 is primarily driven by educational expansion, the joint effect of diverged distribution of education, rate of return to education, and other factors that are not clarified separately included in the “others” of endowment effect, price effect and unobserved skills and abilities residual changes. Rates of return to higher formal education have experienced a large increase relative to their level in 1995, although coefficient estimates of “college or above” and “years of schooling” of the under 35 years old cohort group illustrates lower returns when compared to other groups in Table 4; but the overall rates of return effects have become more severe and extensively dispersed, while the labor market has been characterized by a sharply growing supply of workers with higher education. In the detailed decomposition results, Table A.4 in Appendix shows that the returns to education is approximately 42 percent of the price effect, corresponding to 18.09 percent of the total variation. The education distributional change has primarily altered the distribution of higher education, the proportion of labor force who are at the lower education group has not changed much relative to 1995. Indeed, the distribution of education alone explains more than 50 percent of the endowment effect, which is equivalent to a 22.82 percent of total variation. It is obvious from the table that the change in the distribution of education and its price effect caused 40.91 (22.82+18.09) percent of the widening

of wage inequality. Other discrimination explanatory variables jointly explain 35.75 percent of the endowment effect or equivalently 16.15 of the total variation, and a significant part of price effect as magnitudes of party and ethnic minorities variables narrowed, but offset by regional disparities and gender wage differentials.

## **6. Conclusion and discussion**

As countless writing on wage inequalities show, there are distinct variables that could add to the increment in income inequalities, while, in policy circles, education, education inequality, and educational expansion are regularly considered as the key to address all conceivable social and economic issues.

This paper provides an examination of changes in the Chinese urban wage inequality, from the descriptive statistics, Lorenz curves (increase in curvature), estimation results and decomposition results are immediate evidence of such rising wage disparity among labor market participants in urban China from 1995 to 2013. Since 1999 when an educational expansion took effect, the structure of the education profile encountered a sensational change. In 1995, laborers who had completed college and higher level education represented 8.42 percent of CHIP1995, and college graduates and those who graduated with higher levels of educational achievement comprised very nearly 40 percent in 2013. It is noteworthy that more and more secondary school graduates have the opportunity to enter college or undergraduate university after educational expansion. Female laborers performed extremely better than their male counterparts, the percentage of college and above increased from 29.55 percent to 43.89 percent of the total female labor force. Percentile wage ratios of 90P/50P, 50P/10P, and 90P/10P and wage premiums of college/HS, HS/JS, and college/JS from 1995 to 2013 all have experienced enlarged yields corresponding to increased rates of 19.11 percent, 17.37 percent, and 39.80 percent for

the wage ratios and 11.39 percent, 13.88 percent and 26.85 percent increase for the wage premiums during selected period.

The estimation results indicate the returns to education, which correspond to an extra year of schooling, have improved from 0.0418 to 0.0830. Also we estimates the model by age groups, rate of returns to education seems flatter for CHIP1995, but in 2013, laborers who are 50 years old or older yield higher return when compared to the other age groups, although their education profile is relative low; but competition is also low, income distribution among these laborers are well structured and hard to be changed, and seniority provision protects their position and wages and makes the position of new entrants unfavorable (Thurow 1975). The model includes three discrimination variables, the gap between the ethnic majority and minorities are narrowed, in Table 3, and coefficient estimates are no longer significant even at 10 percent significance level. China Communist Party membership dummy variable is still significant, although the magnitude is getting low, and party variable is not significant for laborers who are 25 to 50 years old, but for age above 50 it yields 0.114 and statistically significant at 1 percent level.

In 1995, gender wage differential gave off an impression of being segmented within the Chinese labor market, with the coefficient estimate yielding 0.114 higher for male workers than the female workers, while in 2013, the magnitude of coefficient estimate extend to 0.288, and both are statistically significant at the one percent level. It is a signal of gender discrimination against the female workers. The age group of 35 years and older appears to be more severe than those who are younger than 35 years old. As more and more (43.89 percent in 2013 and 29.55 percent in 1995 from Table 2) female workers are now entering the skilled level labor market, some of the effects could be offset by relative higher return to education, But there is still an obvious wage differential due to genders. Regional wage

disparity has enlarged, with most of the provinces diverge compared to the reference group, Beijing. Although well-developed coastal provinces like Jiangsu and Guangdong perform better than other provinces, the differences are still noticeable when compared to their levels in 1995. In the regression tables, I included the variance of residuals to capture the inequality of unobservable factors, apparently, it tends to increase from 1995 to 2013.

Up to now, descriptive results and results estimation results suggest a conspicuous wage inequality in the Chinese economy after educational expansion. In section 4, I presented the decomposition method and counterfactual simulations that are required. By utilizing this technique, we are able to break down the change in variance of wages into 1) endowment effect due to structural change; 2) price effect that clarifies the adjustments in coefficient estimates; and 3) the unobserved effect that incorporates the impact of unobserved residuals. Decomposition results show endowment effect and price effect, which together accounts for 70 percent of the change in variance of wages, where the endowment effect is around 50 percent and the price effect pertains to the other 14 percent. By using counterfactual simulation, I isolated the effect of education structural change from the endowment effect, and the latter represents 50.51 percent, that correspond to 22 percent of aggregate change in the fluctuations of the log of the wage from 1995 to 2013. Correspondingly, the education price effect was likewise simulated, and it explains 42.49 percent of the price effect, which represents 18.09 percent of the total variations. Unfortunately, educational expansion does not relieve the pressure of the labor market. On the contrary, it explains a large part of the increment in wage inequality.

Recall the standard human capital equation and its variance transformation from Gregorio and Lee (2002), where the variance of the log of income depends on the variance of schoolings, variance of the rate of return to education, the corresponding means in schoolings and rate of return, and

covariance of educational attainments and rate of return plus a variance of the unobserved error term. In terms of empirical results, all of the terms in this case are presented to be positive in 2013 relative to 1995, thus the change in variance of the log income is increasing. This is so especially, when  $\text{var}(S)$  is huge, that is when the educational inequality of such educational expansion targeting at higher levels education can strongly alter income inequality. Li and Xing (2010) proves the existence of educational inequality using multinomial logit model for those who benefited by the educational expansion policy after high school graduation, and the policy is biased towards certain groups and makes education unequally distributed after 1999. In contrast to the wage competition model, government investments on expanding tertiary education do no help to reduce wage inequality, but only to change the labor characteristics of individuals and increase supply pressures on the labor market which in fact did not narrow the wage differentials between the skilled and unskilled labors. However, from a job competition perspective, the large volume of new entrants may find themselves in a difficult position to retain a higher ranking within the labor queue, and they would be more willing to accept employment that does not match their educational attainments, thereby dragging down the wage level of the unskilled group (Thurow 1975). Thurow argued that the educational investment becomes a defensive tool that only works to defend individual's position in the labor queue, but not necessarily enabling one to really alter his/her real wage income or the distribution of income.

Despite the fact that the outcomes do not seem to support the view that education could easily help to reduce wage inequality, the results can offer suggestions and implications to policymakers about education and wage inequality. More supportive policies can be put forth to protect unfavorable labor market groups, while at the same time having such an expanding supply of tertiary educational investment. In this paper, the author attempts to analyze and break down the impact on wage inequality

of an educational expansion policy, and results presented in this paper only provide evidence that adds to the growing literature on wage inequality in urban China. Along these lines, it does not seek to deny the critical role of education in favoring the growth of the Chinese economy, but it does raise concerns and questions certain claims that this educational expansion can have on the evolution of income distribution in China.

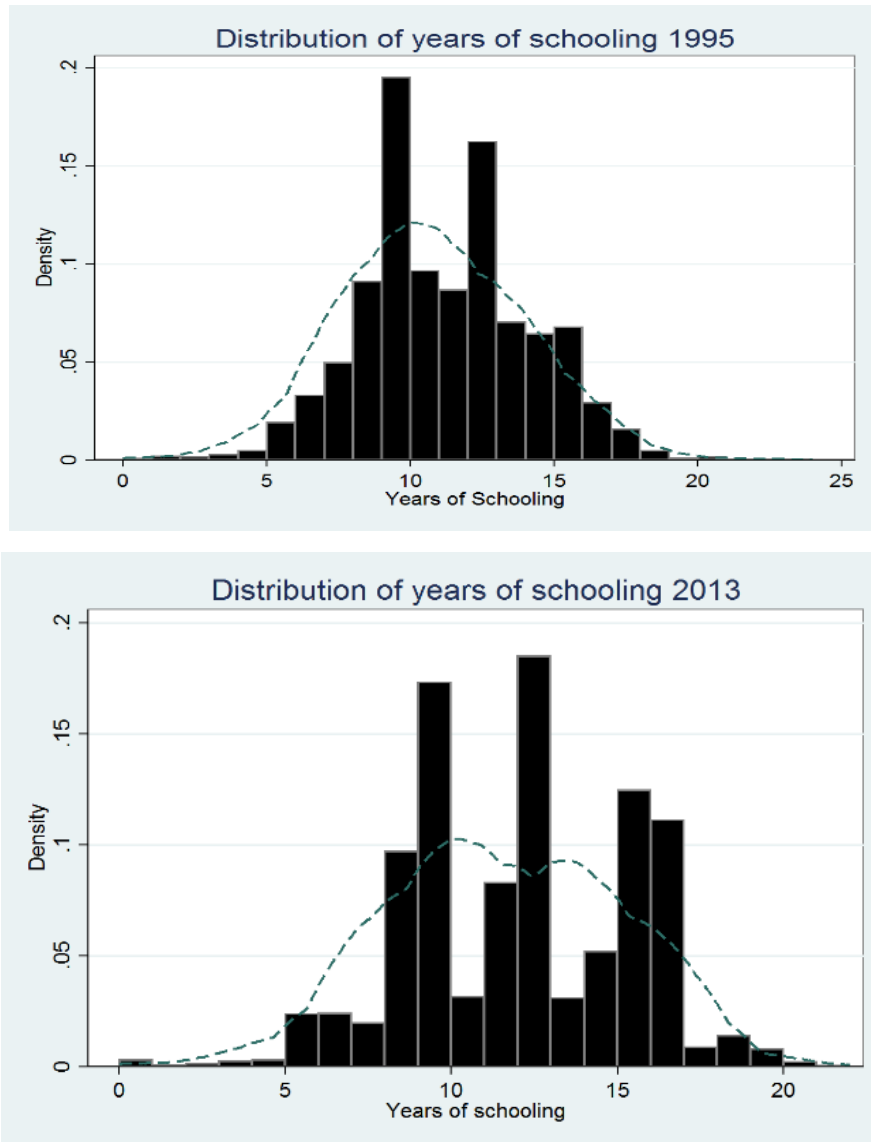
## References

- Appleton, S., Knight, J., Song, L. and Xia, Q. (2009) 'The Economics of Communist Party Membership: The curious case of rising numbers and wage premium during China's Transition.' *Journal of Development Studies*, Vol. 45, No. 2, 256-275
- Battistón, D., C. García-Doménch, and L. Gasparini (2014) 'Could an Increase in Education Raise Income Inequality? Evidence for Latin America.' *Latin American Journal of Economics*, Vol. 51, No. 1, 1-39
- Blinder, Alan S., (1973) 'Wage Discrimination: Reduced Form and Structural Estimates.' *Journal of Human Resources*, Vol.8, 436-55
- Bouillon, C. P., Legovini, A., and Lustig, N. (2003) 'Rising Inequality in Mexico: Household characteristics and regional effects.' *Journal of Development Studies*, 39, 112-133
- Bourguignon, F., F. Ferreira, and N. Lustig (eds) (2005) '*The Microeconomics of Income Distribution Dynamics in East Asia and Latin America* Washington.' DC: Oxford University Press for the World Bank
- Chi, W., R.B. Freeman, and H. Li. (2012) 'Adjusting to Really Big Changes in the Labor Market in China, 1989-2009.' *National Bureau of Economics Research*, Working Paper 17721, January
- Cortez, W. W. (2001) 'What is Behind Increasing Wage Inequality in Mexico?' *World Development* Vol 29, No 11, 1905-1922
- Goldin, C., and L. Katz (2007) 'The Race Between Education and Technology: The Evolution of U.S. Educational Wage Differentials, 1890 to 2005', *NBER Working Paper* No. 12984
- Gregorio J., J. Lee. (2002) 'Education and Income Inequality: New evidence from cross-country data.' *Review of Income Wealth*, Vol. 48, No. 3, 395-416
- Gustafsson, B., S. Li, and T. Sicular, eds. (2008) '*Inequality and Public Policy in China*.' New York: Cambridge University Press
- Knight, J. B., R. H. Sabot. (1983) 'Educational Expansion and the Kuznets Effect.' *The American Economics Review*, Vol. 73, No.5, 1132-1136
- Kuznets, S. (1955) 'Economic Growth and Income Inequality.' *The American Economic Review*, Vol. 45, 1-28
- Li, S., C. Xing (2010) 'China's Higher Education Expansion and its Labor Market Consequences.' *The Institute for Study of Labor (IZA)*, Discussion Paper No. 4974

- Lu, M., G. H. Wan, and Z. Chen (2004) 'Globalization and Regional Income Inequality: Evidence from within China.' *World Institute for Development Economics Research*, Discussion Paper No. 2004/10
- Maurer-Fazio, M. (1999) 'Earnings and Education in China's Transition to a Market Economy: Survey evidence from 1989 to 1992.' *China Economic Review*, 10 (1999) 17-40
- National Bureau of Statistics of China (2014) '*China Statistical Yearbook 2014.*' Beijing: China Statistical Publishing House
- National Bureau of Statistics of China (2016) 'China's Economy Realized a Moderate but Stable and Sound Growth in 2015.' *National Bureau of Statistics of China*, 19 January, 2016
- Oaxaca, Roland, (1973) 'Male-Female Wage Differentials in Urban Labour Markets.' *International Economic Review*, Vol. 14, 673-709
- Song, J., Sicular, T., and Yue, X. (2011) 'Appendix II: The 2002 and 2007 CHIP surveys: Sampling, Weights, and Combining the Urban, Rural, and Migrant Samples', in Li, S., Sato, H., and Sicular, T. *Rising Inequality in China: Challenges to A harmonious Society*. Cambridge University Press.
- Thurow, L. (1972) 'Education and Income Inequality.' *Public Interest*, Vol. 28, 66-81 Sum 72
- Thurow, L. (1975) '*Generating Inequality: mechanisms of distribution in the U.S. economy.*' New York: Basics Books, Inc., Publishers
- Zhou, X. (2014) 'Increasing Returns to Education, Changing Labor Force Structure, and the Rise of Earnings Inequality in Urban China, 1996-2010.' *Social Force*, Vol. 93, No. 2, 429-455

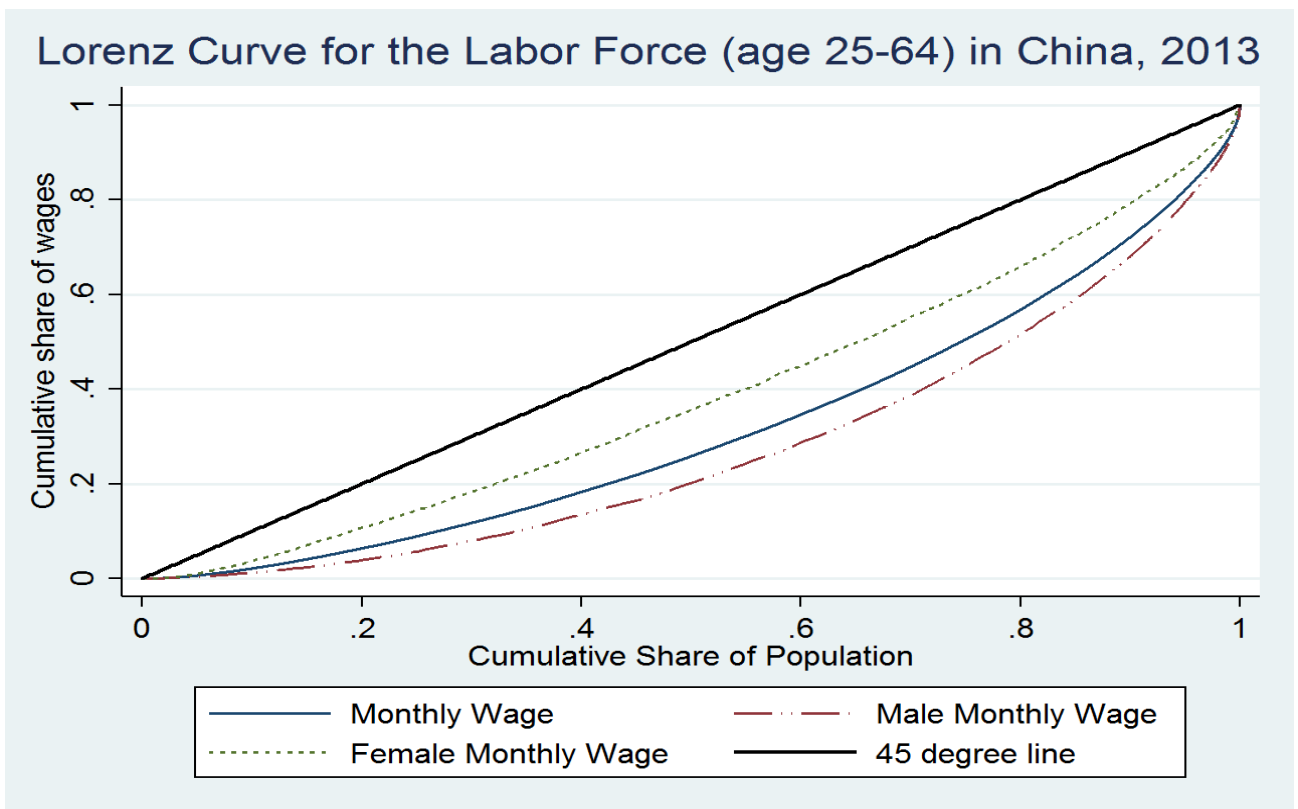
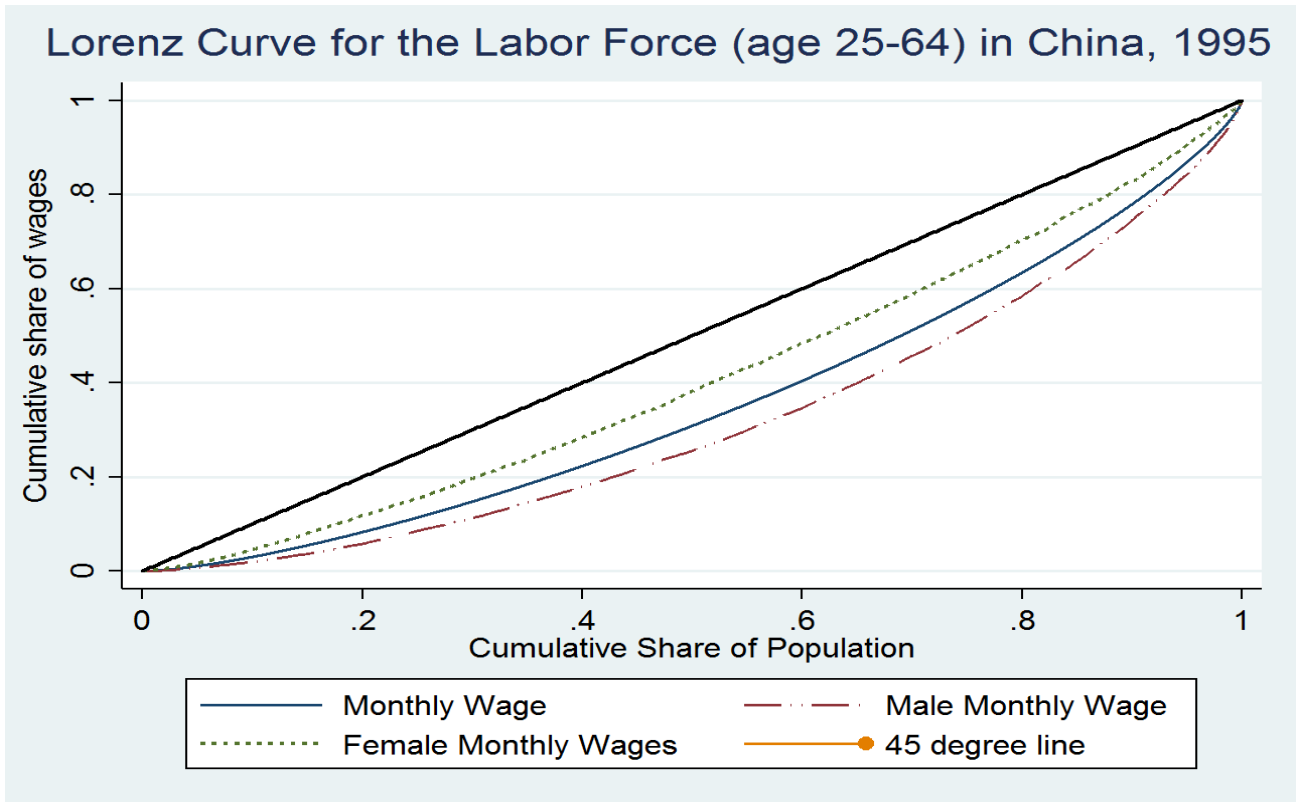
# Appendix

Figure A.1 Distribution of Years of Schooling, 1995 and 2013



Source: author's calculation based on years of schooling from CHIP1995 and CHIP2013

Figure A.2 Lorenz Curve for the Labor Force in China, 1995 and 2013



Source: author's calculation based on monthly wages from CHIP1995 and CHIP2013

Table A.1 Log monthly wage equation, 1995 and 2013

Regression	1995 (1)	1995 (2)	2013 (3)	2013 (4)
Junior	-0.199*** (0.0118)	-	-0.214*** (0.0234)	-
High School	-	-	-	-
college	0.202*** (0.0146)	-	0.356*** (0.0213)	-
Years of Schooling	-	0.0418*** (0.0018)	-	0.0830*** (0.0034)
Experience	0.0373*** (0.0029)	0.0379*** (0.0029)	0.0342*** (0.0032)	0.0334*** (0.0031)
Experience squared	-0.000461*** (0.0001)	-0.000475*** (0.0001)	-0.0007523*** (0.0000)	-0.0006454*** (0.0000)
Han	0.0607*** (0.0231)	0.0592*** (0.0231)	0.0383 (0.0379)	0.0199 (0.0406)
Male	0.114*** (0.0101)	0.114*** (0.0101)	0.302*** (0.0169)	0.288*** (0.0169)
Party	0.084*** (0.0107)	0.0793*** (0.0108)	0.0485 (0.0209)	0.0534 (0.0205)
Constant	5.656*** (0.0419)	5.126*** (0.0481)	6.954*** (0.0457)	6.112*** (0.0518)
Beijing (reference group)	-	-	-	-
Shanxi	-0.4136*** (0.0246)	-0.373*** (0.0245)	-0.5012*** (0.0334)	-0.4392*** (0.0335)
Liaoning	-0.3153*** (0.0238)	-0.3139*** (0.0240)	-0.5001*** (0.0408)	-0.5183*** (0.0404)
Jiangsu	-0.0919*** (0.0221)	-0.0659** (0.0224)	-0.09535*** (0.0308)	-0.0747* (0.0306)
Anhui	-0.3819*** (0.0243)	-0.3756*** (0.0245)	-0.3567*** (0.0363)	-0.2935*** (0.0361)
Shandong	-	-	-0.3576*** (0.0346)	-0.3110*** (0.0344)
Henan	-0.4731*** (0.0244)	-0.4329*** (0.0247)	-0.5041*** (0.0354)	-0.4679*** (0.0352)
Hubei	-0.2768*** (0.0221)	-0.2478*** (0.0223)	-0.3777*** (0.0331)	-0.3368*** (0.0325)
Hunan	-	-	-0.3474*** (0.0346)	-0.3316*** (0.0343)
Guangdong	0.3239*** (0.0257)	0.3675*** (0.0261)	-0.0905*** (0.0312)	-0.0674 (0.0308)
Chongqing	-	-	-0.3685*** (0.0343)	-0.3304*** (0.0343)
Sichuan	-0.2954*** (0.0217)	-0.2713*** (0.0219)	-0.3446*** (0.0356)	-0.3163*** (0.0354)
Yunnan	-0.2114*** (0.0206)	-0.1846*** (0.0207)	-0.4609*** (0.0370)	-0.4381*** (0.0362)
Gansu	-0.4539*** (0.0249)	-0.4082*** (0.0248)	-0.5245*** (0.0342)	-0.4629*** (0.0353)
variance of residuals	0.2584	0.2642	0.4823	0.4766
adj. R-sq	0.279	0.2829	0.255	0.259
Number of observations	10,401	10,273	9,261	9,261

Note: dependent variable is natural log monthly wage, column (1) and (3) are equation (1) and column (2) and (4) are equation (2) for 1995 and 2013. Robust standard errors are in parentheses.

\* p<0.01, \*\* p<0.05, \*\*\* p<0.1

Source: author's calculation based on wage equation (1) and (2) from CHIP1998 and CHIP2013.

Table A.2 Log monthly wage equation by age groups, 1995 and 2013

Regression	CHIP1995			CHIP2013		
	<35 (1)	35-50 (2)	>50 (3)	<35 (4)	35-50 (5)	>50 (6)
Junior	-0.237*** (0.0323)	-0.185*** (0.0129)	-0.175*** (0.0389)	-0.295*** (0.0647)	-0.188*** (0.0301)	-0.148*** (0.0398)
High School	-	-	-	-	-	-
college	0.286*** (0.0308)	0.155*** (0.0176)	0.208*** (0.0343)	0.304*** (0.0434)	0.374*** (0.0277)	0.456*** (0.0453)
Years of Schooling	0.0492*** (0.0045)	0.0372*** (0.0020)	0.0394*** (0.0048)	0.0683*** (0.0054)	0.0714*** (0.0034)	0.0811*** (0.0049)
Experience	0.0547*** (0.0102)	0.0503*** (0.0080)	0.0809*** (0.0125)	0.0547*** (0.0089)	0.0518*** (0.0038)	0.0608*** (0.0063)
Experience squared	-0.00138*** (0.0004)	-0.000708*** (0.0002)	-0.000957*** (0.0002)	-0.00241*** (0.0000)	-0.000687*** (0.0000)	-0.00113*** (0.0005)
Han	0.0581 (0.0415)	0.0844** (0.0298)	0.0435 (0.0654)	0.0453 (0.0759)	0.0671 (0.0529)	-0.1000 (0.1130)
Male	0.118*** (0.0204)	0.101*** (0.0120)	0.161*** (0.0455)	0.184*** (0.0318)	0.349*** (0.0216)	0.336*** (0.0515)
Party	0.0581** (0.0262)	0.0933*** (0.0119)	0.119*** (0.0310)	0.0182 (0.0364)	0.0259 (0.0269)	0.114*** (0.0467)
Constant	5.746*** (0.0870)	5.477*** (0.0993)	4.593*** (0.1947)	7.258*** (0.1221)	7.161*** (0.1767)	6.954*** (0.1792)
Beijing (reference group)	-	-	-	-	-	-
Shanxi	-0.5395*** (0.0516)	-0.3961*** (0.0295)	-0.2573** (0.0806)	-0.5227*** (0.0637)	-0.4631*** (0.0453)	-0.5300*** (0.0803)
Liaoning	-0.4137*** (0.0590)	-0.2976*** (0.0266)	-0.2772*** (0.0752)	-0.6098*** (0.0778)	-0.3671*** (0.0520)	-0.6674*** (0.0974)
Jiangsu	-0.1675*** (0.0520)	-0.0826** (0.0261)	-0.0824 (0.0686)	-0.2318*** (0.0589)	-0.0226 (0.0419)	-0.0804 (0.0721)
Anhui	-0.4698*** (0.0562)	-0.3919*** (0.0294)	-0.2447*** (0.0636)	-0.3515*** (0.0736)	-0.3369*** (0.0471)	-0.3013*** (0.0912)
Shandong	-	-	-	-0.3582*** (0.0643)	-0.2683*** (0.0414)	-0.4682*** (0.1210)
Henan	-0.6288*** (0.0542)	-0.4551*** (0.0299)	-0.2285*** (0.0657)	-0.5560*** (0.0716)	-0.4568*** (0.0458)	-0.5031*** (0.0906)
Hubei	-0.4273*** (0.0538)	-0.2580*** (0.0255)	-0.1607* (0.0647)	-0.4227*** (0.0641)	-0.3286*** (0.0431)	-0.4043*** (0.0906)
Hunan	-	-	-	-0.4339*** (0.0668)	-0.2617*** (0.0451)	-0.4273*** (0.0920)
Guangdong	0.2867*** (0.0659)	0.3308*** (0.0298)	0.324*** (0.0753)	-0.1211** (0.0541)	-0.0393 (0.0421)	-0.1691* (0.0863)
Chongqing	-	-	-	-0.4402*** (0.0627)	-0.3100*** (0.0442)	-0.3596*** (0.0982)
Sichuan	-0.4041*** (0.0537)	-0.3042*** (0.0254)	-0.1035* (0.0614)	-0.3773*** (0.0614)	-0.3349*** (0.0485)	-0.2196** (0.0978)
Yunnan	-0.2793*** (0.0502)	-0.2237*** (0.0236)	-0.1063 (0.0654)	-0.6248*** (0.0777)	-0.3648*** (0.0444)	-0.4302*** (0.1086)
Gansu	-0.6174*** (0.0534)	-0.4385*** (0.0312)	-0.2311*** (0.0645)	-0.4863*** (0.0641)	-0.5253*** (0.0446)	-0.4641*** (0.0914)
Variance of residual	0.3228	0.3184	0.4189	0.4134	0.3736	0.4625
adj. R-sq	0.2183	0.2470	0.3566	0.2113	0.253	0.335
Number of observations	2,878	6,309	1,214	2,386	5,308	1,567

Note: dependent variable is natural log monthly wage. Robust standard errors are in parentheses.

\* p<0.01, \*\* p<0.05, \*\*\* p<0.1

source: author's calculation based on wage equation (1) and (2) between age groups i) age<35, ii) 35≤age≤50 iii) age>50 from CHIP1998 and CHIP2013.

Table A.3 Proportion of different education level by age groups

Education level/ total number of people in age groups			
	age<35	35≤age≤50	age>50
CHIP1995 (College)	0.09	0.06	0.21
CHIP2013	0.56	0.34	0.18
CHIP1995 (High school)	0.69	0.51	0.46
CHIP2013	0.27	0.30	0.33
CHIP1995 (Junior)	0.22	0.43	0.33
CHIP2013	0.17	0.35	0.48

Source: author's calculation

Table A.4 Decomposition results

Decomposition	Variance (LnY1995)	Variance (LnY2013)	Difference	%of total
Total effect	0.35237936	0.788367	0.4359876	100%
	$f(X_{1995}, \beta_{2013}, \varepsilon_{2013})$	$f(X_{2013}, \beta_{2013}, \varepsilon_{2013})$	Difference	% of total
Endowment	0.591388	0.388387	0.196999	45.18%
	$f(X_{1995}, \beta_{1995}, \varepsilon_{2013})$	$f(X_{1995}, \beta_{2013}, \varepsilon_{2013})$	Difference	% of total
Price	0.405722	0.591388	0.185666	42.59%
	$f(X_{1995}, \beta_{1995}, \varepsilon_{1995})$	$f(X_{1995}, \beta_{1995}, \varepsilon_{2013})$	Difference	% of total
Unobserved	0.35237936	0.405722	0.0533426	12.23%
Endowment effect		of 0.196999	of the total 0.4359876	
Education		50.51%		22.82%
Experience		13.74%		6.21%
Others		35.75%		16.15%
Price effect		0.171687	of the total 0.4468267	
Education		42.49%		18.09%
Experience		-8.94%		-3.81%
Variance (LnY1995)		Variance (LnY1995), $\beta_{edu}=2013$	Difference	
0.35238		0.431271833	0.0788918	
Variance (LnY1995)		Variance (LnY1995), $\beta_{exp}=2013$	Difference	
0.35238		0.335788	-0.016592	
Variance (LnY1995)		Variance (LnY1995), $edu=2013$	Difference	
0.35238		0.451879332	0.0994993	
Variance (LnY1995)		Variance (LnY1995), $exp=2013$	Difference	
0.35238		0.379452234	0.0270722	

Source: Author's calculation on CHIP1995 and CHIP2013