DISCUSSION PAPER SERIES

IZA DP No. 4919

# Economic Reform, Education Expansion, and Earnings Inequality for Urban Males in China, 1988-2007

Xin Meng
Kailing Shen
Sen Xue

April 2010

Forschungsinstitut
zur Zukunft der Arbeit
Institute for the Study
of Labor

# Economic Reform, Education Expansion, and Earnings Inequality for Urban Males in China, 1988-2007

**Xin Meng**
*Australia National University*
*and IZA*

**Kailing Shen**
*WISE, Xiamen University*
*and IZA*

**Sen Xue**
*WISE, Xiamen University*

Discussion Paper No. 4919
April 2010

IZA

P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0
Fax: +49-228-3894-180
E-mail: iza@iza.org

IZA Discussion Paper No. 4919
April 2010

# ABSTRACT

## Economic Reform, Education Expansion, and Earnings Inequality for Urban Males in China, 1988-2007

In the past 20 years the average real earnings of Chinese urban male workers have increased by 350 per cent. Accompanying this unprecedented growth is a considerable increase in earnings inequality. Between 1988 and 2007 the variance of log earnings increased from 0.27 to 0.48, a 78 per cent increase. Using a unique set of repeated cross-sectional data this paper examines the causes of this increase in earnings inequality. We find that the major changes occurred in the 1990s when the labour market moved from a centrally planned system to a market oriented system. The decomposition exercise conducted in the paper identifies the factor that drives the significant increase in the earnings variance in the 1990s to be an increase in the within-education-experience cell residual variances. Such an increase may be explained mainly by the increase in the price of unobserved skills. When an economy shifts from an administratively determined wage system to a market-oriented one, rewards to both observed and unobserved skills increase. The turn of the century saw a slowing down of the reward to both the observed and unobserved skills, due largely to the college expansion program that occurred at the end of the 1990s.

JEL Classification: J31, P2, P3

Keywords: earnings inequality, China

Corresponding author:

Xin Meng
Division of Economics
Research School of Economics
Australian National University
Canberra 0200
Australia
E-mail: Xin.Meng@anu.edu.au

# 1 Introduction

Over the past 20 years China has had unprecedented economic growth. In 1988, China's per capita GDP averaged $377, and by 2007 this value had increased almost sevenfold to $2604 (United Nation data series). This same period has also seen considerable institutional changes in an economy which switched from a centrally-planned to a market-oriented system; a large scale rural to urban migration; and a huge expansion in high education enrollment. In 1988 around 98 per cent of the urban labour force worked either in the state or the collective sectors, only 10 per cent of the urban labour force had a college or above degree, and rural to urban migration was almost non-existent. By 2007, employment in the state or collective sectors had dropped by 36 percentage points to 63 per cent, labour force with tertiary and above education[1] increased by 30 percentage points, while around 140 million rural labour force (accounting for one third of the urban labour force) moved to cities to work. We would expect institutional and structural changes of this magnitude should have a significant impact on the distribution of the earnings of local urban workers.

This paper sets out to document and investigate the impact of these changes on the changes in urban male workers'[2] earnings dispersion. Using a unique set of repeated cross-sectional data we show that the last 20 years has seen the variance of the log male annual earnings in urban China almost doubling. Unlike the earnings growth, however, the changes in earnings dispersion have been more erratic. There was hardly any change in earnings inequality in the period between the late 1980s and the early 1990s, but the variance of the log earnings increased from 0.23 in 1992 to 0.47 in 2001. Since then, the increase in earnings inequality has been slow, up until 2007.

This paper contributes to the literature by linking the changes in earnings inequality with the measurable structural and institutional changes occurring during the same period. We show

[1] In the Chinese case, this includes individuals with 3-year college and above education.

[2] 'Urban' in here has a special meaning related to the household registration system in China. Since the early 1950s, China implemented a household registration system, whereby individuals born in a city are given an 'urban' household registration, and those who were born in countryside are given a 'rural' household registration. Such a system has not changed for the past 60 years in that individuals who were born to parents with 'rural' household registration are still given a 'rural' household registration even though he/she might work in cities. In this paper we use the term 'urban workers' to refer to workers who have 'urban' household registration.

that part of this significant increase in earnings inequality can be explained by the switching of the urban labour force from the state/collective to the private sectors and part of it is due to the increase in the return to education, and the reduction of the return to seniority. However, for the most part, the changes are related to the significant changes in within-group residual variances.

The second contribution of this paper is that we examine the factors which might have affected the changes in within-group residual variances. The literature has always assumed that the change in the within-group residual variances is driven by the change in the price for unobserved skills while the distribution of unobserved skills is held constant over time (see, for example, Juhn, Murphy, and Pierce, 1993 and Lemieux, 2002 and 2006). We find that in China it is impossible to argue that the distribution of unobservable skills has not changed, even within the skill cells. In fact we find that although the changes in the within-group residual variances in the 1990s might be closely related to the increase in the price for unobserved skills, the slowing down of the increase in within-group residual inequality in the 2000s may be mainly affected by a significant increase in the supply of young educated workers. This not only suppressed the price for observed and unobserved skills, but also changed the distribution of within-group unobserved skills.

The paper is structured as follows. The next section presents background information on factors affecting earnings inequality, such as changes in labour market institutions and supply of and demand for skilled and unskilled labour during this period. Section 3 introduces the data used and describes changes in earnings inequality during the period. Section 4 discusses the methodology. Section 5 investigates the changing pattern of earnings inequality and the main contributing factors to the changes in inequality. Section 6 examines the causes ofs the change in the contribution of the within-group residual variances. Conclusions are given in Section 7.

## 2 Background

Under the planned economic regime, urban workers had lifetime employment whereby their jobs were assigned by the state, firms were not allowed to fire workers, and wage levels were

centrally determined. The market oriented economic reform which began in the agricultural sector at the end of the 1970s took a long time to spread to the urban sector. By the end of the 1980s, although some progress in marketization was made in the goods market, the urban labour market reform lagged behind. Neither job mobility nor wage determination had changed much from the planned economic regime, except that enterprises had the right to retain a proportion of their profits which were often used to pay bonuses to workers (Korzec, 1992; White, 1988; Takahara, 1992 and Meng, 2000).

The early 1990s saw the speed-up of marketization in the goods market and an increase in the dispersion of profitability within the state sector. Studies find that in the early 1990s wage variations depended heavily on firms' profitability in the state sector (Meng, 2000).

It took the financial collapse of many state-owned enterprises to initiate the real changes in the urban labour market. In the mid 1990s, more than 40 per cent of the state-owned enterprises were making losses. To vitalize the economy, the government finally made it legal for the state enterprises to become bankrupt and to fire unwanted workers. It was not until 1997 that a policy to 'keep the 1000 large scale state-owned enterprises and let go the small and medium loss-making firms' was implemented and in response there was a significant shift in the labour force from the state to the non-state sector. In 1997-98 around 12 million urban state sector workers were laid off (Meng, 1998).

It was against this background that more and more workers shifted to private sector employment. Panel A of Figure 1 presents the total share of state sector employment for the urban workers and for those who had five or less years of work experience. It shows that since 1997-98, the proportion of workers employed in the state sector has been reducing rapidly, especially for new entrants to the labour market. In 1997, around 75 per cent of new entrants were employed in the state sector and by 2007 this figure had reduced to 40 per cent.

While most urban people were still working in the state/collective sectors, private sector employment grew quickly. Since the early 1990s, China has embarked on an export-oriented economic growth strategy. The growth of the FDI and joint venture firms increased demand for unskilled workers in cities. This increase was met by the large scale inflow of rural migrants to urban centres. According to the data available to us, in 1994 there were around 37 million rural

migrant workers in cities, and by 2007 this had increased to 135 million (See Panel C of Figure 1). The most significant inflow occurred around the end of the 1990s. The increase in migrant labour supply was accompanied by a high proportion of the urban labour force exiting from the labour market, and a shift of urban workers from unskilled production to service and clerical occupations (panel B of Figure 1). Panel C of Figure 1 shows the increase in the working age (16-60 years) of the urban non-student population who were not in the labour market, together with the number of rural migrants working in cities for the period 1988 to 2007. The two trends increase almost hand-in-hand.

Another important change in China, which might have a significant impact on urban wage inequality, is the expansion of higher education. Education levels of urban Chinese have been affected by significant political changes. During the ten year Cultural Revolution (1966-1976), schools were closed for many years. Universities did not begin normal student recruitment until 1977. As a result, an entire cohort of youth missed schooling for various numbers of years (Meng and Gregory, 2002 and 2005a). After the Cultural Revolution, universities reopened for the public and anybody who missed the opportunity to go to universities because of the Cultural Revolution was allowed to sit the National College Entrance Examinations and those who were successful were able to go to university regardless of their age.[3] As a result, in the 1980s and early 1990s the proportion of the labour force with a tertiary degree increased significantly. Throughout the 1990s the proportion of the work force with tertiary education increased only slightly. Then, in 1999 the government decided to increase tertiary (three-year college and above) enrolments by as much as 40 per cent (panel D of Figure 1). This university enrollment expansion significantly increased the supply of the skilled workers from the level observed early this century. Panel E of Figure 1 depicts the change in the share of employed males with different levels of education. At the beginning of this period (1988-2007), more than 50 per cent of the male workers had education levels at or below junior high school and by the end of the period this ratio had dropped to around 20 per cent. A reverse trend is found for college or university degree holders. In 1988 a little more than 10 per cent of male workers had a tertiary education and by the end of the period the ratio had increased to 40 per cent.

[3] College entrance in China normally has an age restriction.

Finally, because of the introduction of the One-Child Policy at the end of the 1970s, the urban population has aged significantly (panel F of Figure 1). The proportion of young workers with one to ten years of experience dropped from around 30 per cent of the total workforce at the beginning of the period, to just below 20 per cent at the end. The proportion of those with 21 to 30 years of experience increased from 23 per cent to 34 per cent.

We now analyse how these changes have affected the earnings distribution in the past 20 years.

# 3 Data and the general picture of earnings inequality

## 3.1 Data

The data used in this study come from the Urban Household Income and Expenditure Survey (UHIES) 1988-2007. The survey was conducted by the National Bureau of Statistics annually since 1956, and was resumed in 1980 after its suspension during the Cultural Revolution—1966-1976 (Fang et al. 2002). The electronic data are available from 1986. The survey samples households with Urban Household Registration for every province in the nation. We use the data from 16 of the 31 provinces including Beijing, Shanxi, Liaoning, Heilongjiang, Shanghai, Jiangsu, Anhui, Jiangxi, Shandong, Henan, Hubei, Guangdong, Sichuan, Chongqing, Yunan, and Gansu.

This is the only data set in China which goes back 20 years to record annual changes in earnings during this extraordinary growth period. The survey was conducted for the purpose of monitoring income and expenditure changes for households whose household registrations (Hukou) are located in urban areas. It uses a diary record to collect individual earnings, other forms of income, household income and expenditure related data. Basic individual demographic and human capital variables are also available. The data have been widely used to analyse income inequality, poverty, rates of return to education, and household savings in China (see, for example, Han, Wailes, and Cramer, 1995; Fang et al., 2002; Gibson, Huang, and Rozelle, 2003; Meng, Gregory, and Wang, 2005b; Zhang, Zhao, Park, and Song, 2005; Meng, Gregory

and Wan, 2007; and Chamon and Prasad, 2008).[4]

The UHIES questionnaire changed four times in 1988, 1992, 1996, and 2002. The important change for this paper occurred in 1988. In 1986 to 1987, the only income information collected at the individual level is the monthly standard wage. From 1988 onwards, the survey collected individual work related earnings including standard wages, floating wages, bonuses, subsidies, and other work related wages. Because the earnings measured in the 1986 and 1987 survey are only a subset of earnings, these years are excluded from the analysis.

Another data related issue worth mentioning is the change in sampling procedure since 2002. Before then the survey only sampled households with urban household registrations. Since 2002 the sampling covers all households with a residential address in an urban city where the survey is conducted. However, the majority of rural-urban migrants either live on the periphery of cities where rents are cheaper or in dormitories or workplaces such as construction sites. Only a very limited number of households with rural household registration are included in the survey. For example, in the 2003 survey there are around 2 per cent of individuals with rural household registration in the sample. To keep consistency in the sampling population, individuals with rural household registration in the 2002-2007 surveys are excluded. In addition to the change in sampling coverage, sample size has also more than doubled since 2002.

As this study is on male earnings inequality, the sample includes male individuals aged 20 to 60 who are working and have a positive wage at the survey year.[5] Excluding missing values, for each of the 20 years the final sample size ranges from 7400 to 26,900. Table 1 presents the summary statistics of the data. It shows that over the 20 years, the mean age of the male workers increased slightly from 39 to 42 years. Similarly, actual and potential work experience[6] all increased slightly. Further, the average years of schooling increased from 11 years to almost 13 years, while the proportion of workers with tertiary and above education increased from 16 per cent at the beginning of the period to 41 per cent at the end. The most significant change

[4] The survey design and sampling procedure have been discussed in great detail in the above mentioned papers.

[5] In urban China the majority of labour force work works full time. For example, using CHIPS 1995 and 2002 data where days and hours worked variables are available, we found that in 1995, 97% of the working population in urban China worked 7 or more hours a day, while 99% worked 5 or more days a week. In 2002 the proportion who worked 7 or more hours a day is 95%, and 97% worked 5 or more days a week.

[6] Potential work experience is calculated as: age minus years of schooling minus 7.

is the increase in male annual real earnings[7], which increased from 4,379 yuan in 1988 to 19,520 yuan in 2007, an increase of 346%.

### 3.2 General picture of earnings inequality

The above discussion focuses mainly on the mean values. As discussed in Section 2, the significant increase in mean earnings over the last 20 years was accompanied by changes in labour market institutions and the structural change in labour demand and supply. The latter brought about considerable shifts in the distribution of earnings.

Table 2 presents summary measures of earnings inequality during the 20 year period, including the Gini coefficient, the ratio of the 90th to the 10th percentiles of real annual earnings, and the variance of the logarithmic of real annual earnings. The table shows that overall earnings inequality increased significantly, by any measure. The Gini coefficient increased from 0.23 to 0.37, an increase of 58 per cent. The ratio of the 90th to the 10th percentiles was 2.9 times in 1988 and increased to 5.4 times in 2007, an increase of 87 per cent. Similarly, the variance of the log real annual earnings increased by 76 per cent over 20 years. Such an increase is much higher than that found for the US where over the period of 26 years between 1963 to 1989, there is a 72 per cent increase in variance of log weekly earnings (Juhn, Murphy, and Pierce, 1993).

Figure 2 shows the earnings inequality from different aspects. All of them convey the same story of a fast increase in earnings inequality. Below we pay particular attention to panels C, D and E.

Panel C presents the change in real earnings for the 10th, 50th, and the 90th percentiles. We observe that in the first few years the earnings of the three groups went up almost equally. Between the early to the end of the 1990s the 90th percentile went up much more than the 50th and the 10th percentile groups. Over this period, the bottom income group (the 10th percentile) hardly changed from its early 1990s level. This pattern coincides well with the fact that during this period drastic state enterprise reform led to a significant drop in earnings for the loss making firms. After the year 2000, the earnings of low income workers started to grow

[7] Real earnings are calculated using nominal earnings deflated by provincial level CPI which is set at 100 for 1995.

again, but the rate of change was slower than that of the 90th percentile group.

Panel D of Figure 2 shows the ratio of the final years (2005–2007) and the base years earnings (1988–90) for each percentile to indicate the percentage changes over 20 years for the entire distribution. To reduce the measurement error problem, we take the ratio of the average of the last three years (2005–2007) and that of the first three years (1988–1990). The scatter plot indicates the actual changes while the solid line shows the linear prediction. The graph reveals that over the 20 year period the real earnings at the bottom more than doubled, while at the top they increased by five or more times.

The final panel (Panel E of Figure 2) illustrates the same ratios for different periods. There, we see different patterns over time. The period from the late 1980s to the early 1990s had very small changes in earnings inequality, while the decade beginning in 1990 has the largest increase in inequality. In this period there is a real earnings drop for the very low earnings group, but a more than 60 per cent increase for the very high earnings group. The years between the late 1990s and early 2000s also witnessed large increases in inequality but that was mainly driven by the larger increase in earnings at the top of the distribution. Earnings increased most dramatically relative to other periods between the early to mid 2000s, but the increase is almost equally distributed and earnings inequality did not change significantly.

The above descriptive analysis focuses mainly on the overall inequality. As significant institutional and structural changes have occurred in the Chinese urban labour market, the changes in earnings inequality may have affected different groups differentially, as indicated in Section 2. Figure 3 presents the variance of the log real earnings by level of work experience, education, occupation, and sector of employment. The graphs indicate earnings inequality for within-groups. Panel A shows that earnings inequality is more serious among younger workers than it is among the older ones. However, since the late 1990s, there has not been a further widening of earnings inequality among younger workers, whereas for the older workers the inequality continued to increase until 2004. This pattern of wider inequality among the less experienced is unusual from the point of view of western labour markets, where within-group inequality is usually higher among the experienced (Juhn, Murphy, and Pierce, 1993 and Lemieux, 2006). This difference, perhaps, is the result of the significant labour market transition from a planned to a market-

oriented economy. It is the young workers who are more likely to enter into the market-oriented private sector, where earnings dispersion has been much wider than in the public sector (see Panel C of Figure 3). Indeed, even at the beginning of the period (1988), 27 per cent of the group with 1–10 years of experience were not working in the state sector, while for the older group the ratio was 10 per cent. Over time this gap increased. By the end of the period 52 per cent of the young group were not working in the state sector, whereas for those with more than 20 years of experience the proportion is 35 per cent. Another reason for the young to have a larger within-group earnings variance may be related to the steepness of the age-earnings profile during the first ten years of their career.

Panel B of Figure 3 depicts the variance of log real earnings for individuals with tertiary education (three-year college and above, labeled 'educated' hereafter) and those with junior high school and below education (labeled 'less-educated' hereafter). Unlike many western countries, earnings inequality is much higher among the less-educated group. This, perhaps, is related to the fact that during the Cultural Revolution (1966–1976) entire cohorts of youth were not able to go to school and hence were forced into the less educated group even though their innate ability might otherwise allow them to enter the educated group had they been given the opportunity to do so. Thus, within the less-educated group the ability distribution should be wider than that of the educated group. Between the 1980s and the 1990s, as China moved from a planned to a market-oriented economy, ability was gradually being rewarded and hence we observe the sharp increase in within-group earnings inequality during this period for both the educated and less-educated. However, as the Cultural Revolution cohorts gradually retire and more educated young people enter the labour market, especially after the university expansion in the late 1990s, the inequality of earnings within the less educated group begins to drop, while within the educated group inequality continues to increase. This changing pattern of the ability composition within the educated and less-educated groups will remain an important issue throughout the paper.

Panel C of Figure 3 investigates the earnings dispersion within the state (state and collective sectors) and non-state sectors. As expected, the earnings dispersion within the non-state sector is much larger than in the state sector. From the mid-1990s onwards, however, the inequality

within the non-state sector dropped significantly for reasons unclear to us at this point. Note that the sample size for the non-state sector is very small for the data between 1988 and 1993 and relatively small for the period 1994 to 1996. Thus, it is possible that the extremely high variance during the period of 1991–1996 is due to sample noise and should not be taken seriously. Within the state sector the earnings inequality started off from a fairly low level and increased continuously during the 1990s. Since the early 2000s the within state sector earnings variance has kept constant.

The last panel (Panel D) examines different patten of inequality for different occupation groups. Service/clerical workers and production workers exhibit larger within-group earnings inequality than the highly educated professional and managerial groups. The decline of earnings inequality among production workers from 1999 onwards may be related to two factors: First, the large scale inflow of the rural migrant workers crowded out many urban local workers who exited from the labour market. We might assume that the crowding out effect mainly affected the bottom of the ability distribution and hence narrowed down the earnings inequality among urban local production workers. Second, the large scale university expansion from the late 1990s has allowed lower ability workers to move up to better skilled jobs, which in turn reduced the earnings distribution at the bottom. The latter effect may be confirmed by the same pattern of decline in inequality for the production workers and for the less-educated group (see Panel B for junior high school and below group).

The general inequality picture presented here indicates that there has been a considerable increase in earnings inequality, occurring mainly in the 1990s. The changes in inequality among different groups of the labour force, however, is not the same.

# 4 Methodology

The increase in earnings inequality is often decomposed into the change in the price of observable characteristics, the change in the distribution of observable characteristics, and the change in the residual, which is normally labelled as 'the price of unobserved skills' under certain assumption (See, for example, Juhn, Merphy, and Pierce, 1993; DiNardo, Fortin). We follow Lemieux (2002

and 2006) to decompose the change in variance of log earnings over time. Consider the following earnings equation:

$$y_{it} = x_{it}\beta_t + \epsilon_{it}, \tag{1}$$

where $y_{it}$ is the log of the annual earnings of individual $i$ at year $t$; $x_{it}$ is a vector of observable characteristics (we include education and labour market experience for the Baseline Model and education, experience, and state sector employment for the Extended Model); $\beta_t$ is a vector of return to observable characteristics; and $\epsilon_{it}$ is the error term. Lemieux (2002) shows that the variance of the log earnings can be written as:

$$V_{it} = Var(y_{it}) = \sum_i \omega_{jt}(x_{it}\widehat{\beta}_t - \overline{x}_t\widehat{\beta}_t)^2 + \sum_i \omega_{jt}u_{it}^2 = \sum_j \theta_{jt}(\overline{y}_{jt} - \overline{y}_t)^2 + \sum_j \theta_{jt}\sigma_{jt}^2, \tag{2}$$

where $\omega$ is the sample weight, $\widehat{\beta}$ is the vector of OLS estimates of $\beta$, and $u$ is the residual from the OLS regression of equation (1). In the second part of the equation (2), the subscript $j$ indicates the number of skill cells in the sample (for example, if $x$ is a vector of experience and education categories, $j$ will be the number of experience-education cells in the sample); $\sigma_{jt}^2$ is the within-cell residual variance, and $\theta_j$ is the proportion of observations in cell $j$ of the sample (in the case where $x$ are all dummy variables, $\theta$ will be $\overline{x}$).

To decompose the change in the variance of earnings between year $t$ and year $s$, Lemieux (2002) proposes two counterfactual variances. The first is to replace $\widehat{\beta}_t$ in Equation (2) with $\widehat{\beta}_s$ which is obtained by estimating Equation (1) with period $s$ data. Thus, the first counterfactual variance, denoted as $V^a$, is defined as:

$$V_{it}^a = \sum_i \omega_{jt}(x_{it}\widehat{\beta}_s - \overline{x}_t\widehat{\beta}_s)^2 + \sum_i \omega_{jt}u_{it}^2 = \sum_j \theta_{jt}(\overline{y}_{js} - \overline{y}_t^a)^2 + \sum_j \theta_{jt}\sigma_{jt}^2. \tag{3}$$

The second counterfactual variance, denoted as $V^b$, is to replace $\theta_{jt}$ in Equation (3) by $\theta_{js}$, which is the actual proportion of observation in cell $j$ for period $s$:

$$V_{it}^{b} = \sum_{i} \omega_{js}(x_{is}\widehat{\beta}_{s} - \overline{x}_{s}\widehat{\beta}_{s})^{2} + \sum_{i} \omega_{js} u_{it}^{2} = \sum_{j} \theta_{js}(\overline{y}_{js} - \overline{y}_{s}^{a})^{2} + \sum_{j} \theta_{js}\sigma_{jt}^{2}. \tag{4}$$

Using the two counterfactual variances, $V^a$ and $V^b$, we can decompose the change in variance of earnings between periods $t$ and $s$ into three components: 1. a portion which is due to the change in the return to the observables ($\Delta\widehat{\beta}$); 2. a portion which is due to the change in the composition of the sample ($\Delta\theta_j$); and 3. a portion which is due to the change in the within-cell residual variance ($\Delta\sigma_j$). Thus, $V_t - V_s$ may be written as:

$$V_t - V_s = \underbrace{(V_t - V_t^a)}_{Observable\ price} + \underbrace{(V_t^a - V_t^b)}_{Composition} + \underbrace{(V_t^b - V_s)}_{Within\ cell\ variance}. \tag{5}$$

Lemieux (2006) further investigates the residual variance component and writes it as follows:

$$Var(u_{it}) = p_t^2 \sum_{j} \theta_{jt}\sigma_{jt}^2, \tag{6}$$

where $p_t$ is price for unobservables. He then assumes that the distribution of unobservable skills among workers within each skill cell $j$ is stable over time, thus, $\sigma_{jt}^2 = \sigma_{js}^2 = \sigma_j^2$. Because of this, Equation (6) can be re-written as

$$Var(u_{it}) = \sum_{j} \theta_{jt} V_{jt}, \tag{7}$$

where $V_{jt}$ varies only if the price for unobservables varies over time.

The assumption that the distribution of the unobservable skills within skill cell is constant over time is unrealistic for China. One example may be worth mentioning. As discussed in Section 2, the significant expansion in college enrolment in 1999 enabled a large number of less capable people to obtain degrees. Thus, within the college cell the unobserved skill distribution should widen significantly and, at the same time, the distribution of the unobservable skill distribution for the less educated group should narrow. Hence, the last term in Equation (5) not only represents the change in price for the unobservables, but also the change in their distribution.

# 5 Decomposition of earnings inequality

## 5.1 Results from the earnings equations

We first estimate the earnings equation (Equation (1)) for each of the 20 years between 1988 and 2007. To reduce measurement error problems, we exclude the top and bottom 1 per cent of the sample. Two specifications of the earnings equation are used. The baseline model controls for four categories of education (three-year college and above, technical school, senior high school, and junior high school and below), four categories of years of work experience (1-10 years, 11-20 years, 21-30 years, and 31 and above years) and provincial fixed effects. This specification is consistent with that used in most U.S. studies on wage inequality. The inclusion of provincial dummy variables is necessary to control for significant spatial price variations (Brandt and Holz, 2006 and Gong and Meng, 2008).[8] In addition, we also extend the baseline model with an additional dummy variable indicating whether an individual is working in the state sector or not because the switch between the state and non-state sector is probably an important factor in the change in earnings inequality.

Tables 3 and 4 present the results for the Baseline and the Extended Models, respectively. In both cases the omitted category is 1-10 years experience and junior high school and below. The results in Table 3 show that there seems to be a very significant return to years of experience. The longer the experience, the higher the return throughout the 20 year period. However, relative to the group with 1-10 years of experience, the change in the return to experience varies significantly for different experience groups over different periods. In the late 1980s and early 1990s those with 11-20 years of experience earned around 35 per cent more than those with 1-10 years of experience. Since the early 1990s the return to experience for this group has reduced to around 25-28 per cent. For those with more work experience (21-30 years and above 30 years of experience), the reduction in relative returns has been more dramatic and continuous. For individuals with 21-30 years of experience the returns reduced from 48 per cent in 1988 to 34 per cent in 2007, while for those with more than 30 years of experience the return

[8] We assume that provincial variations are fully captured by the changes in $\beta$s and $X$s and will have no effect either on residual variance distribution across groups or within groups.

drops from 55 per cent to 35 per cent. The most significant reduction occurred in the period between the early 1990s and the early 2000s when the economy experienced the most significant deregulation. During the planned economic era, wages were administratively determined and seniority was one of the most important considerations in the earnings determination process. The market oriented economic reform has gradually abolished the seniority determined wage system and hence, at the end of the reform period the return to work experience has started to reflect labour productivity (Meng and Kidd, 1997 and Meng, 2000).

While the return to experience has been reducing, the return to education has been increasing. Compared to the less-educated group (junior high school and below), the senior high, technical school, and three-year college and above groups in 1988 earned 2, 7, and 13 per cent more. By 2007, the same ratios increased to 17, 27, and 56 per cent, respectively.

Panels A and B of Figure 4 present these changes visually. The most important message evident in Panel A is the continuous decline in the rate of returns to experience, especially for the groups with 21 and above years of work experience. Panel B shows that the most significant increase in returns to tertiary education (college and above) occurred during the 1990s. By the early 2000s this trend had flattened out. This, perhaps, is related closely to the college expansion policy. The patterns for technical and senior high schools are similar, but to a lesser extent.

The Extended Model (Table 4) presents a similar picture of the change in rates of return to education and work experience. The additional variable, employment in state- or collectively-owned enterprises, has had dramatic changes in its payoffs. During the late 1980s and early 1990s, employees in the state or collective sectors were paid higher earnings than their counterpars in other sectors. These payoffs had disappeared in the 1990s, and then in the mid-1990s the return became negative as a large proportion of the state- or collective-owned firms were making losses. It was not until the early 2000s that the payoff to state and collective employment was again positive (see Panel C of Figure 4).

Another way to gauge the contribution of each set of variables included in the regressions when explaining the total earnings variation is to derive a bounded range of the contributions using partial and marginal $R^2$s for each set (see Dickens and Katz, 1987). We plot the upper

bound (the partial $R^2$s) of the contributions.[9] These are presented in Panel D of Figure 4. Some striking features are worth noting. First, in the late 1980s, experience alone explained around 25 per cent of the earnings variances and the effect reduced dramatically over time. By 2007 it only explained less than 3 per cent of the earnings variation, suggesting the disappearing of a seniority-driven wage determination system. Second, in the 1990s, provincial dummy variables explained around 20 per cent of the earnings variations and it was not until the late 1990s that this effect decreased and stayed at around 10 per cent of the earnings variation for the rest of the period. Finally, education did not become an important factor until after the mid-1990s.

Another issue worth mentioning is the extent to which earnings variations can be explained by all the observed characteristics. The estimated earnings regression results show a decline of the adjusted $R^2$s. In the late 1980s and early 1990s the adjusted $R^2$s are around 0.33 to 0.38. From the mid-1990s the adjusted $R^2$s start to decline, and by the end of the period, the observable characteristics can only explain 23 to 25 percent of the earnings variations. Thus, over time, the residual component of the earnings variations has increased significantly.

### 5.2 Decomposing the total earnings inequality

We apply Lemieux's (2002) decomposition method in order to understand what generated the large increase in earnings inequality over the last 20 years. Table 5 presents the change in variance of earnings between 1988 and 2007, as well as for four different time periods within this 20-year period, (i.e. 1988-1992, 1992-1998, 1998-2003, and 2003-2007). The top panel of the table uses the Baseline Model with 1988 as the base year. The first column presents the change in variance between 1988 and 2007. During the 20 year period, the total variance of log earnings increased by 0.21, which is a 76 per cent increase to its 1988 level of 0.27. Of the total change in variance, 91 per cent is due to the increase in the sum of the within-group residual variances. The composition effect (that is the change in the proportions of people in different education-experience cells) contributes to a narrowing of the earnings inequality by 13 per cent. The change in returns to observable characteristics contributes to the increase in the earnings

[9] We also calculated the lower bound of the impact of each set of variables on earnings variation by adding each set of the variables to the existing sets of variables to see their marginal contribution to the $R^2$s. These results are similar to those presented in Panel D of Figure 4 and are available upon request from the authors.

variation. These two effects, however, almost cancel each other out.

The next four columns split the period into four sub-periods. In the 1988-1992 period the urban economy operated under the old planned economy system and earnings inequality reduced slightly. Of the small change observed, the composition effect and the changes in the within-group residual variances contribute almost equally to the narrowing in the earnings inequality.

Urban economic reform in China took place mainly in the 1990s. During this period, many state-owned factories were making losses and more-able workers switched to the private sector while less-able workers exited from the labour market. It was during this period that the earnings inequality widened the most. Over the seven years between 1992 and 1998, the variance of the log earnings increased from 0.16 to 0.32, which accounted for more than three quarters of the total increase in the variance of log earnings over the entire 20 years. Of the total change of 0.16 variance, 88 per cent can be attributed to the increase in the within-group residual variances.

The 1998-2003 period continue to see the increase in variance of the log earnings (by 0.08), which is the second largest change among the four sub-periods. Once again the increase in the within-group residual variances explains the most of the change in the earnings inequality.

During the 2003–2007 period the earnings inequality declined and it was mainly the change in the within-group residual variances which contributed to this decline. The negative contribution of the within-group residual variances may, to a large extent, be related to the significant inflow of tertiary graduates to the labour market from 2002 onwards. This increase in the supply of skilled labour should have played an important role in the slowing down of earnings inequality. We will come back to this point later in this section.

The above results suggest that during the 1990s when earnings inequality increased most significantly the major contributing factor to inequality is the increase in within-group residual inequality, whereas both the late 1980s to the early 1990s period and the 2000s period experienced a reduction in the earnings inequality. Panel A of Figure 5 presents the decomposition results for each of the 20 years, where 1988 is used as the base year. Thus, the changes in the return to observable characteristics are predicted using every year's regression coefficients, but 1988 $X$s. Similarly, the changes in composition effects and the changes in within-group residual

variances are both calculated relative to 1988. The trend presented in this figure confirms the dominant role which the changes in the within-group variance played in explaining the increase in earnings inequality during the 1990s and the slowing down of the increase during the 2000s.

The second panel of Table 5 presents the same decomposition results, but this time using 2007 as the base year. As all decompositions suffer from the index number problem, this panel is presented so to allow us to gauge the bounds of the effects. The trends indicated here (and in Panel B of Figure 5) are largely consistent with those obtained using 1988 as the base year.

The last panel of Table 5 shows the results from the Extended Model, which includes an additional dummy variable indicating whether an individual is employed in the state/collective sector or the non-state sector. The decomposition uses 2007 as the base year.[10] Relative to the results presented in the second panel of this table, we find that including the state/collective employment dummy variable increases the composition effect and reduces the within-group residual variance effect significantly. When not controlling for state/collective employment, the composition effect contributes a mere 4 per cent reduction in the change in earnings variations. Once the state/collective employment is controlled for, the composition effect becomes positive 23 per cent, suggesting that the shift from state to non-state employment contributes to 27 per cent (4 per cent plus 23 per cent) of the total increase in the earnings variations. Similarly, the effect of within-group residual variances reduced from contributing to 88 per cent of the total change to 65 per cent of the total change. This is quite a significant effect, indicating that the difference in earnings inequality between the state and non-state sectors is an important part of the story of the increase in earnings inequality during this period. We also find that the major changes occurred between 1998 and 2003, and this coincides with the significant reduction in the state/collective employment as presented in Panel A of Figure 1 in Section 2.

Even though including state/collective employment increases the contribution of the composition effect, the general dominance of the within-group residual variance over the period does not change. For the entire period, it still accounted for 65 per cent of the changes in the variance of the earnings, and for the period 1993–1998 it accounted for 87 per cent of the total

[10] In 1988, 98 per cent of the sample were employed in either the state or collective sectors. Thus, if we use 1988 as the base year, many education/experience/sector-of-employment cells will be empty, and this makes the decomposition impossible.

change. This can also be seen from Panel C of Figure 5.

# 6 What caused the changing pattern of the within-group residual variances?

The above analysis indicates that the most significant increase in earnings inequality for urban male workers occurred in the 1990s and the main contributing factor is the increase in the sum of the within-group residual variances. From the early 2000s the increase in earnings inequality has stopped and so has the increase in within-group residual variances. The question naturally arises as to what are the driving forces behind the changing pattern of the residual variances over the past 20 years? To answer this question we first examine whether the change in the dispersion of within-group residual variances has the same trend for different skill groups. Figure 6 presents the changes in residual log earnings variances by education and experience cells. While Panels A and B of Figure 6 show the different trends in the residual variance for different experience groups within the educated (college and above) and less educated (Junior high and below) groups, respectively, Panels C and D present the changes for different education levels within the 1–10 (labeled as 'young') and 21–30 (labeled as 'old') years of experience groups.

The figure shows that the significant increase in the within-group residual variances between the early 1990s and the early 2000s is common for both young and old, and educated and less-educated groups. However, since the early 2000s the growth of the within-group residual variances has plateaued, but the speed of change varied across different skill groups. While in the educated group there is a similar slowing down of the increase in the within-group variances for the young and the old groups, the rate of the slowing down in the less-educated group for the young seems to be much faster than that for the old. The latter generated a significant narrowing of the gap in the within-group residual variances between the old and young in the less-educated group (Panel B) and for the young across different education categories (Panel C).

### 6.1 Causes of the increase in the within-group residual variances in the 1990s

What could be the reason for the considerable increase in the residual variance from the early 1990s to the early 2000s? In the literature changes in within-group residual variances are often fully attributed to changes in price for unobserved skills while the within-group distribution of unobserved skills is assumed to be held constant over time (Juhn, Murphy and Pierce, 1993 and Lemieux, 2002 and 2006).[11] This is quite a reasonable assumption if over time the proportion of the labour force across different skill cells does not change very much. As indicated in the Background Section, over the past 20 years the skill levels of the urban male labour force has changed significantly, but in the 1990s (especially between 1994-99) the changes have been moderate (see Panels E of Figure 1). This leads us to believe that it is unlikely that the distribution of the unobserved skill within each group has changed much for this period, and hence the increase in the within group residual variances is more likely to be driven by the increase in the price for unobserved skills. After all, it is reasonable to assume that the observed and unobserved skills are highly correlated, and hence, the increase in returns to education in the 1990s could also be a sign of the increase in returns to unobserved skills. The question, however, is what might have generated the increase in the returns to unobserved skills.

There could be two possible causes. One is the increase in demand for unobserved skills, of which 'skill biased technical change' is often cited as the driving force (Acemoglu, 2002). The other possible cause is the changes in the wage determination system due to the shift from a planned to a market economy. In the planned economic system, wages were administratively determined and wage distribution across different skill groups was very compressed (see, for example, Meng, 2000; Benjamin, Brandt, Giles, and Wang, 2009). The process of marketization has increased the returns to education and other productivity related observed and unobserved skills.

A recent study on the causes of the increase in the returns to education for urban local workers in the 1990s has attributed most of the increase in returns to education to skill-biased technical change (Liu, Park, and Zhao, 2010). However, the empirical method used in their

[11] Lemieux (2002 and 2006) also discussed the measurement error as a potential reason. As it is less likely to be an important contributing factor for China, we do not discuss this issue here.

study may not be able to disentangle the two possible causes discussed above. Following Katz and Murphy (1992) the authors estimate a version of the following equation:

$$\ln(\frac{W_h}{W_l})_t = \alpha T + \beta \ln(\frac{N_h}{N_l})_t + \epsilon_t, \tag{8}$$

where $(\frac{W_h}{W_l})$ is relative earnings of skilled and unskilled workers, and $\frac{N_h}{N_l}$ is the relative employment of the skilled and unskilled workers. The basic idea of this method is that the relative supply of skilled labour increases and moves the employment ratio down the downward slopping demand curve over time and this effect will be captured by $\beta$. If there is any (skill-biased) technical change occurring over time, it shifts the demand curve (upward) and this effect will be captured by the coefficient on the time trend ($\alpha$). Given that the elasticity of substitution is greater than 1, a positive $\alpha$ will imply a skill-biased technical change. However, in the Chinese case a positive $\alpha$ is also consistent with the fact that the institutional change towards a market economy has increased wages of skilled workers relative to unskilled workers. Thus, simply obtaining a positive $\alpha$ cannot identify whether the significant increase in within group residual variances is caused by the increase in demand for skilled workers or by institutional change.[12]

In fact, we are more inclined to believe that the latter may have played a more important role. As discussed earlier, China experienced its most significant labour market reform in the 1990s. During this period we observe a considerable drop in the returns to experience (seniority) and increase in the returns to education (Figure 4), reflecting the switching regime of wage determination. Furthermore, if skill-biased technical change is the main driving force, we might observe a continuation or even a speeding-up of the increase in within-group residual variances in the 2000s as this is an even faster economic growth period. Instead, the 2000s has seen a slowing down of the increase in within-group residual variances.

[12] Another potential issue related to using Katz and Murphy (1992) style method to examine the effect of skill-biased technical change is related to lack of suitable data. As discussed in Liu, Park, and Zhao (2010) apart from the census 1990 and 2000 and inter-census 1995 and 2005, almost all the other data recording the urban labour market in China exclude rural-urban migrants, which accounted for around 30% of the urban labour force in 2005.

### 6.2 Why did the increase in within-group residual variances plateau at a different rate for different skill groups in the 2000s?

Given our conjecture that the significant increase in within-group residual variances in the 1990s is mainly caused by institutional changes in wage setting and given that this process was virtually completed by the early 2000s, we would expect that the increase in within-group residual variances should stop from the early 2000s. This, indeed, is the case (see Figure 6).

However, what is surprising is that the slowing down of the increase in within-group residual variances occurs at a different rate for different skill groups. Within the college group the gap in residual variances between the young and old remained constant over the entire period, but for the less-educated group there was a narrowing of the gap in the residual variances between the young and the old. There was also a narrowing of the gap between educated and less-educated groups for those with 1 to 10 years of experience. In other words, the rate of slowing down is much faster for the young than for the old and within the young it is much faster for the less-educated than for the educated. This pattern is more obvious if we reproduce the two graphs from Panels C and D of Figure 6, excluding the technical and senior high school groups (see Figure 8). For the young group (1-10 years of experience), the gap between residual variances for the educated and the less-educated groups narrowed significantly in the 2000s. This pattern, however, is not observed for the old group (21–30 years of experience) where, in the 2000s, the variance between the more and less educated groups has widened. Why is this the case?

We believe that this difference, to a large extent, is related to the change in the within-group distribution of unobserved skills. Even though the literature often assumes that the within-group distribution of unobserved skills is held constant over time (Juhn, Murphy and Pierce, 1993 and Lemieux, 2002 and 2006), this assumption may not be valid in our setting.[13]

One possible cause for over-time changes in the distribution of within-group unobserved ability is the significant changes in education attainment. As discussed in the background section, the significant education expansion occurred in 1999 when there was a sudden increase in college and university enrolment of more than 40 per cent. Since then the speed of the

[13] In fact Lemieux (2002) has discussed in great detail how the change in within-group distribution of unobserved skill may affect the changes in within-group residual variances.

expansion has not slowed (see Panel D of Figure 1). Panel E of Figure 1 shows that in 1988, 16 per cent of the total male labour force had a tertiary qualification and by 2007 this ratio had increased to above 40 per cent. While the proportion of technical and senior high school graduates did not change much over the same period, the ratio of the work force with junior high school or below education reduced from over 48 per cent in 1988 to less than 24 per cent in 2007.

How should the increase in college enrolment affect the within-group distribution of unobserved skills? The following example may illustrate this clearly. Assume that the unobserved skill distributions for the two periods for the population are constant, and that the total population is divided into two groups: the educated and the less-educated. Assume that in period one, only 10 per cent of the population is in the educated group, and in period two this increases to 40 per cent. In this situation, the within-group distribution of unobserved skills for the educated group should be wider in the second period than in the first period, while for the less-educated group it should be narrower in the second period than that in the first period. This is because only relatively able people can pass the college entry exam and receive an education at a college or higher level. When there is a significant expansion of tertiary enrolment, the less-able people will be admitted to tertiary education, and hence widen the ability distribution within the educated group and narrow the ability distribution within the less-educated group.

This distributional change of unobserved skills for the educated and less-educated groups is illustrated in Figure 8, where the normal distribution curve represents the distribution of unobserved skills for the entire population for both period 1 ($t$) and period 2 ($t+1$). The two grey bars indicate the division between the educated and less-educated groups in period 1 ($\text{high}_t$) and period 2 ($\text{high}_{t+1}$). As society decides to educate more people it lowers the 'standard' required to be in the educated group (the grey bar moves to the left). As a result, the distribution of unobserved skills within the educated group widens from the distance 'educated t' to the distance 'educated $t+1$', while that of the less-educated group narrows by the same amount (from 'less-educated $t$' to 'less-educated $t+1$').

Of course, the variance of unobserved ability within the educated and less-educated groups not only depends on the "distance" but also on the shape of the distribution. To ensure that the

above claims are valid for our data, we carry out the following simulations. First we generate four different random samples of 5,000 observations each with a normal distribution. The first sample has mean zero and standard deviation equal to one. For the next three samples, the means and standard deviations are set the same as our 1988, 1997, and 2006 samples. We then rank these observations within each sample and calculate the variance for each of the four full samples and then for the samples excluding an additional 100 observations at a time from the top of the distribution to the bottom (i.e. the first sample has 5,000 observation, the second excludes the top 100 observations, the third exclude additional 100 observations, until the fiftieth sample, which has 100 observation of the lowest value). This gives us 50 variances for each experiment for each sample. If the "distance" is positively related with the variances given the mean and the standard deviation of the distribution, we should observe that each new variance calculated with 100 less observations should be smaller than the previous one. Our results for all four samples (each with 1,000 experiments) are presented in Appendix A, which shows that in all cases our original conclusion that the "distance" is positively correlated with the variance is true given the shape of our chosen distributions.[14]

Our simple example shows that when the proportion of the labour force with tertiary education doubles while proportion of labour force with junior high school education reduces by half, it is not possible to assume that the within-group distribution of unobserved ability is held constant over time and all the changes in the within-group residual variances are due to the change in price for the unobserved ability. Based on the timing of the nationwide education expansion, we would expect that in the 2000s, the change in the within-group distribution of unobserved skills would play an increasing role in explaining the change in the within-group residual variance. In addition, we would expect that the distributional change would affect the young more than the old and that it should increase the within group residual variance for the educated and reduce it for the less-educated.

Indeed, our empirical results are rather consistent with these predictions. To show this more clearly we take difference-in-differences of Panels A and B of Figure 7 (the difference between the educated and less-educated in the young group minus the difference between the educated

[14] More details of the simulations are available upon request from the authors.

and less-educated in the old group). If the difference between the educated and less-educated in the young group mimics that in the old group we should see the diff-in-diff plot hovering around the zero line. Figure 9 presents this diff-in-diffs plot and it shows that for the late 1980s and most of the 1990s the diff-in-diffs values do not deviate from zero much, but since the 2000s, the difference of within-group residual variances between the educated and less-educated for the young group is widened significantly relative to that for the old group.

Although the pattern of changes in within-group variances among different groups presented in Figure 7 fits in with our predictions, there are some details which do not fit exactly. In particular, although the left panel of Figure 7 indicates a reduction in within-group residual variance for the less-educated in the young group in the 2000s, it does not show a similar degree of increase in the within-group variance for the educated group, as we predicted. Our conjecture for this inconsistency is that, perhaps, the significant increase in the supply of educated labour force has not only suppressed the price for observable skills but also that for unobservable skills. Thus, the increased 'distance' in distribution increases the residual variance for this group but the supressed price for unobservables offset the distribution effect. This may be confirmed by the fact that without significant changes in the distribution of the unobservable skills for the educated in the old group, the within-group residual variance for this group has also declined. Thus, the decrease in the price for unobservable skills offsets the effect of the increase in the distribution of the unobservable skills for the young educated group. Had there not been a slowing down of the rewards to the observable and unobservable skills, we would have observed an increase in within-group residual variances for the educated young group.

# 7 Conclusions

In this paper we examine the change in male earnings inequality in urban China during a period of unprecedented economic growth and significant institutional changes. We find that over the twenty year period, earnings inequality increased by almost 60 per cent when inequality is measured by the Gini coefficient and 76 per cent when it is measured by the variance of log real earnings. Such a significant increase has not been seen in many other countries.

Our analysis indicates that while the changes in prices for observable skills and the changes in composition of the male labour force explained some of the changes, more than 80 per cent of the increases in earnings inequality during this period was due to the increase in the within-group residual inequality.

Our analysis suggests that during the 1990s the increases in within-group residual variances are mainly due to the increases in returns to unobservable skills. As the economy transformed from a system where wages were administratively determined to a market system, the reward to both observed and unobserved skills increased.

In the 2000s, however, we note that the trend in within-group residual variances changes significantly. This is a result of a combination of the changes in the distribution of the within-group unobserved skills and a reduction in the reward to the unobservable (as well as observable) skills due to the over supply of the labour force with tertiary qualification. As a result of the significant expansion of college enrolments from the late 1990s, the within-group unobserved skill distribution widened in the educated group and narrowed in the less-educated group. We found that this change in the distribution of within-group unobserved skills could, perhaps, explain a large part of the slowing down of the changes in within-group residual variances.

## References


Acemoglu, D., 2002, "Technical change, inequality, and the labor market", *Journal of Economic Literature*, 40(1), pp.7-72.

Benjamin, D., Brandt, L., Giles, J., and Wang, S., 2008, "Income inequality during China's economic transition", in Loren Brandt and Tom Rawski (eds) *China's Great Economic Transformation*, Cambridge University Press.

Brandt, Loren and Holz, Carsten A., 2006, "Spatial price differences in China: estimates and implications", *Economic Development and Cultural Change*, 55 (1), pp. 43-86

Chamon, M. and Prasad, E., 2008, "Why are saving rates of urban households in China rising?", *NBER Working Paper Series* No. 14546.

Dickens, W. T. and Katz, L. F., 1987, "Inter-industry wage differences and industry characteristics" in K. Lang and J. S. Leonard (eds) *Unemployment and the Structure of labour Markets*, Oxford: Basil Blackwell Inc., pp.48-89.

DiNardo, J., Fortin, N.M., and Lemieux, T., 1996, "labour market institutions and the distribution of wages, 1973-1992: a semiparametric approach", *Econometrica*, 64(5), pp.1001-1044.

Han, T., Wailes, E.J., and Cramer, G.L., 1995, "Rural and urban data collection in the People's Republic of China", in *The China Market Data and Information Systems, Proceedings of WCC-101 Symposium*, Washington, D.C.

Fang, C., Zhang, X., and Fan, S., 2002, "Emergence of urban poverty and inequality in China: evidence from household survey", *China Economic Review*, 13, pp.430-443.

Gibson, J., Huang, J., and Rozelle, S., 2003, "Improving estimates of inequality and poverty from urban China's household income and expenditure survey", *Review of Income and Wealth*, 49(1), pp.53-68.

Gong, C.H. and Meng, X., 2008, "Regional price differences in urban China 1986-2001: Estimation and implication", *IZA Discussion Paper* No. 3621.

Juhn, C., Murphy, K.M., and Pierce, B., 1993, "Wage inequality and the rise of returns to skill", *Journal of Political Economy*, 101(3), pp.410-442.

Korzec, Michael, 1992, *labour and the Failure of Reform in China*, London: MacMillan Press

Ltd, 1992.

Lemieux, T., 2002, "Decomposing changes in wage distributions: a unified approach", *Canadian Journal of Economics*, 35(4), pp.646-688.

Lemieux, T., 2006, "Increasing residual wage inequality: composition effects, noisy data, or rising demand for skill?", *American Economic Review*, 96(3), pp.461-498.

Lie, X., Park, A., and Zhao, Y., 2010, "Explaining rising returns to education in urban China in the 1990s", Unpublished manuscript.

Meng, X. and Kidd, M., 1997, "Wage determination in China's state sector in the 1980s", *Journal of Comparative Economics*, 25(3), pp.403-421.

Meng, X., 2000, *Labour Market Reform in China*, Cambridge: Cambridge University Press.

Meng, X. and Gregory, R. G., 2002, "The impact of interrupted education on subsequent educational attainment: A cost of the Chinese Cultural Revolution", *Economic Development and Cultural Change*, 50(4), pp.935-959.

Meng, X., Gregory, R.G. and Wang, Y., 2005, "Poverty, inequality, and growth in urban China, 1986-2000", *Journal of Comparative Economics*, 33(4), pp 710-729.

Meng, X. and Gregory, R. G., 2005, "The impact of interrupted education on earnings: the educational cost of the Chinese Cultural Revolution" *IZA Discussion Papers,* No. 2548.

Meng, X., Gregory, R. and Wan, G., 2007, "China urban poverty and its contributing factors, 1986-2000" *Review of Income and Wealth*, 53(1), pp.167-189.

Takahara, Akio, 1992, *The Politics of Wage Policy in Post-Revolutionary China*, London: MacMillan Press Ltd.

White, Gordon, 1988. "State and market in China's labour reforms", *Journal of Development Studies*, 24, 4:180-202.

Zhang, J., Zhao, Y., Park, A., and Song, X., 2005, "Economic returns to schooling in urban China, 1988 to 2001", *Journal of Comparative Economics*, 33, pp.730-752.

**Table 1:** Summary Statistics, 1988-2007

| | 1988 | | 1989 | | 1990 | | 1991 | | 1992 | | 1993 | | 1994 | | 1995 | | 1996 | | 1997 | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Mean | SD | Mean | SD | Mean | SD | Mean | SD | Mean | SD | Mean | SD | Mean | SD | Mean | SD | Mean | SD | Mean | SD |
| Age (year) | 39.03 | 10.41 | 39.23 | 10.31 | 39.44 | 10.34 | 38.96 | 9.99 | 39.13 | 10.03 | 39.35 | 9.97 | 39.16 | 9.85 | 39.51 | 9.69 | 39.73 | 9.59 | 39.82 | 9.37 |
| Years of schooling (year) | 10.95 | 2.89 | 11.11 | 2.92 | 11.23 | 2.87 | 11.39 | 2.92 | 11.52 | 2.73 | 11.58 | 2.67 | 11.72 | 2.68 | 11.77 | 2.65 | 11.78 | 2.62 | 11.77 | 2.63 |
| % with 3 year college & above | 0.16 | | 0.18 | | 0.19 | | 0.21 | | 0.24 | | 0.24 | | 0.26 | | 0.26 | | 0.26 | | 0.26 | |
| Actual work experience (year) | 19.40 | 10.44 | 19.96 | 10.35 | 20.29 | 10.38 | 19.85 | 10.09 | 19.95 | 10.15 | 20.20 | 10.04 | 20.11 | 10.00 | 20.46 | 9.82 | 20.68 | 9.78 | 20.71 | 9.65 |
| Potential work experience (year) | 21.03 | 10.92 | 21.07 | 10.90 | 21.19 | 10.88 | 20.56 | 10.63 | 20.59 | 10.57 | 20.76 | 10.47 | 20.44 | 10.40 | 20.74 | 10.22 | 20.94 | 10.19 | 21.05 | 10.04 |
| log real annual earnings | 8.27 | 0.52 | 8.25 | 0.50 | 8.32 | 0.48 | 8.37 | 0.47 | 8.51 | 0.48 | 8.56 | 0.54 | 8.63 | 0.61 | 8.67 | 0.60 | 8.69 | 0.61 | 8.72 | 0.64 |
| Real annual earnings (yuan) | 4379 | 1978 | 4271 | 1985 | 4537 | 2117 | 4765 | 2369 | 5520 | 3108 | 5991 | 3647 | 6632 | 4349 | 6849 | 4334 | 7052 | 4821 | 7390 | 5150 |
| No. of observations | 7,890 | | 7,087 | | 7,470 | | 7,587 | | 9,845 | | 9,477 | | 9,443 | | 9,357 | | 9,414 | | 9,376 | |
| | 1998 | | 1999 | | 2000 | | 2001 | | 2002 | | 2003 | | 2004 | | 2005 | | 2006 | | 2007 | |
| | Mean | SD | Mean | SD | Mean | SD | Mean | SD | Mean | SD | Mean | SD | Mean | SD | Mean | SD | Mean | SD | Mean | SD |
| Age (year) | 40.09 | 9.26 | 40.42 | 9.12 | 40.31 | 9.30 | 40.98 | 9.13 | 41.55 | 9.00 | 41.81 | 8.93 | 42.19 | 8.93 | 42.01 | 9.08 | 42.38 | 9.01 | 42.36 | 9.16 |
| Years of schooling (year) | 11.88 | 2.60 | 12.00 | 2.58 | 12.04 | 2.67 | 12.01 | 2.63 | 12.35 | 2.67 | 12.42 | 2.65 | 12.48 | 2.67 | 12.53 | 2.71 | 12.64 | 2.69 | 12.75 | 2.68 |
| % with 3 year college & above | 0.28 | | 0.29 | | 0.31 | | 0.30 | | 0.36 | 0.48 | 0.37 | 0.48 | 0.38 | 0.48 | 0.39 | 0.49 | 0.40 | 0.49 | 0.42 | 0.49 |
| Actual work experience (year) | 20.96 | 9.60 | 21.32 | 9.53 | 20.82 | 9.89 | 21.42 | 9.83 | 21.06 | 9.71 | 21.31 | 9.72 | 21.60 | 9.79 | 21.24 | 9.97 | 21.56 | 9.94 | 21.96 | 10.18 |
| Potential work experience (year) | 21.20 | 9.95 | 21.43 | 9.83 | 21.27 | 10.09 | 21.97 | 9.94 | 22.19 | 9.93 | 22.39 | 9.85 | 22.70 | 9.86 | 22.47 | 10.01 | 22.73 | 9.94 | 22.60 | 10.13 |
| log real annual earnings | 8.76 | 0.66 | 8.84 | 0.66 | 8.90 | 0.70 | 8.98 | 0.69 | 9.14 | 0.66 | 9.20 | 0.69 | 9.31 | 0.68 | 9.40 | 0.69 | 9.49 | 0.69 | 9.64 | 0.69 |
| Real annual earnings (yuan) | 7725 | 5611 | 8318 | 5329 | 9161 | 6768 | 9875 | 7137 | 11320 | 7932 | 12298 | 9203 | 13740 | 10769 | 15196 | 12098 | 16723 | 13473 | 19520 | 18132 |
| No. of observations | 9,186 | | 9,043 | | 8,841 | | 8,668 | | 21,581 | | 24,411 | | 25,785 | | 26,701 | | 26,967 | | 24,736 | |

Source: Authors' own calculation using the UHIES data 1988-2007.

**Table 2:** Summary measures of earnings inequality, 1988-2007

| | Gini(real annual earnings) | Ratio of 90th to 10th percentile of real annual earnings | Variance(log real annual earnings) |
|---|---|---|---|
| 1988 | 0.232 | 2.879 | 0.275 |
| 1989 | 0.238 | 2.932 | 0.247 |
| 1990 | 0.231 | 2.899 | 0.230 |
| 1991 | 0.234 | 2.739 | 0.222 |
| 1992 | 0.242 | 2.775 | 0.233 |
| 1993 | 0.275 | 3.246 | 0.297 |
| 1994 | 0.309 | 3.917 | 0.367 |
| 1995 | 0.300 | 3.745 | 0.356 |
| 1996 | 0.311 | 3.853 | 0.374 |
| 1997 | 0.320 | 4.011 | 0.413 |
| 1998 | 0.327 | 4.216 | 0.430 |
| 1999 | 0.316 | 4.292 | 0.436 |
| 2000 | 0.341 | 4.881 | 0.490 |
| 2001 | 0.340 | 4.769 | 0.472 |
| 2002 | 0.323 | 4.539 | 0.420 |
| 2003 | 0.354 | 5.362 | 0.541 |
| 2004 | 0.362 | 5.238 | 0.506 |
| 2005 | 0.369 | 5.530 | 0.524 |
| 2006 | 0.367 | 5.493 | 0.502 |
| 2007 | 0.368 | 5.384 | 0.483 |

Source: Authors' own calculation using the UHIES data 1988-2007.

**Table 3:** Selected regression results from the baseline model, 1988-2007

| | 1988 | 1989 | 1990 | 1991 | 1992 | 1993 | 1994 | 1995 | 1996 | 1997 |
|---|---|---|---|---|---|---|---|---|---|---|
| Experience 11-20 years | 0.364*** | 0.346*** | 0.361*** | 0.286*** | 0.267*** | 0.287*** | 0.270*** | 0.278*** | 0.279*** | 0.264*** |
| | [0.0108] | [0.0119] | [0.0109] | [0.0107] | [0.00949] | [0.0114] | [0.0128] | [0.0131] | [0.0136] | [0.0146] |
| Experience 21-30 years | 0.484*** | 0.467*** | 0.493*** | 0.422*** | 0.402*** | 0.436*** | 0.439*** | 0.415*** | 0.404*** | 0.390*** |
| | [0.0117] | [0.0124] | [0.0110] | [0.0108] | [0.00947] | [0.0112] | [0.0130] | [0.0132] | [0.0135] | [0.0143] |
| Experience >30 years | 0.548*** | 0.539*** | 0.577*** | 0.494*** | 0.475*** | 0.496*** | 0.507*** | 0.484*** | 0.490*** | 0.452*** |
| | [0.0130] | [0.0135] | [0.0121] | [0.0124] | [0.0110] | [0.0131] | [0.0152] | [0.0153] | [0.0157] | [0.0170] |
| College | 0.127*** | 0.146*** | 0.164*** | 0.172*** | 0.196*** | 0.239*** | 0.306*** | 0.253*** | 0.269*** | 0.256*** |
| | [0.0115] | [0.0118] | [0.0104] | [0.00997] | [0.00860] | [0.0101] | [0.0116] | [0.0116] | [0.0119] | [0.0126] |
| Technical school | 0.0683*** | 0.0821*** | 0.122*** | 0.125*** | 0.128*** | 0.149*** | 0.192*** | 0.164*** | 0.170*** | 0.152*** |
| | [0.0130] | [0.0139] | [0.0120] | [0.0116] | [0.0103] | [0.0121] | [0.0141] | [0.0143] | [0.0146] | [0.0159] |
| Senior high school | 0.0210** | 0.016 | 0.0146 | 0.0413*** | 0.0470*** | 0.0698*** | 0.0889*** | 0.0678*** | 0.0834*** | 0.0895*** |
| | [0.0103] | [0.0108] | [0.00955] | [0.00945] | [0.00856] | [0.0101] | [0.0119] | [0.0117] | [0.0119] | [0.0126] |
| Provincial dummies | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Observations | 7729 | 6941 | 7317 | 7435 | 9648 | 9288 | 9254 | 9170 | 9220 | 9184 |
| R-squared | 0.35 | 0.33 | 0.39 | 0.38 | 0.38 | 0.37 | 0.36 | 0.33 | 0.33 | 0.32 |
| | 1998 | 1999 | 2000 | 2001 | 2002 | 2003 | 2004 | 2005 | 2006 | 2007 |
| Experience 11-20 years | 0.275*** | 0.251*** | 0.278*** | 0.278*** | 0.254*** | 0.292*** | 0.308*** | 0.268*** | 0.288*** | 0.281*** |
| | [0.0155] | [0.0162] | [0.0171] | [0.0181] | [0.0110] | [0.0113] | [0.0108] | [0.0106] | [0.0107] | [0.0111] |
| Experience 21-30 years | 0.397*** | 0.362*** | 0.386*** | 0.383*** | 0.328*** | 0.352*** | 0.343*** | 0.317*** | 0.325*** | 0.342*** |
| | [0.0151] | [0.0158] | [0.0167] | [0.0175] | [0.0107] | [0.0109] | [0.0105] | [0.0104] | [0.0106] | [0.0109] |
| Experience >30 years | 0.459*** | 0.420*** | 0.437*** | 0.411*** | 0.387*** | 0.403*** | 0.391*** | 0.343*** | 0.339*** | 0.345*** |
| | [0.0185] | [0.0180] | [0.0188] | [0.0190] | [0.0120] | [0.0121] | [0.0117] | [0.0117] | [0.0115] | [0.0116] |
| College | 0.324*** | 0.387*** | 0.448*** | 0.462*** | 0.486*** | 0.521*** | 0.522*** | 0.534*** | 0.511*** | 0.555*** |
| | [0.0133] | [0.0136] | [0.0149] | [0.0151] | [0.00938] | [0.00964] | [0.00920] | [0.00923] | [0.00922] | [0.00950] |
| Technical school | 0.189*** | 0.236*** | 0.276*** | 0.249*** | 0.255*** | 0.274*** | 0.276*** | 0.270*** | 0.266*** | 0.274*** |
| | [0.0166] | [0.0172] | [0.0194] | [0.0196] | [0.0121] | [0.0125] | [0.0120] | [0.0124] | [0.0124] | [0.0130] |
| Senior high school | 0.104*** | 0.142*** | 0.144*** | 0.179*** | 0.162*** | 0.161*** | 0.163*** | 0.171*** | 0.151*** | 0.170*** |
| | [0.0134] | [0.0139] | [0.0155] | [0.0154] | [0.00999] | [0.0102] | [0.00986] | [0.0101] | [0.0100] | [0.0103] |
| Provincial dummies | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Observations | 9002 | 8861 | 8662 | 8490 | 20926 | 23667 | 25002 | 25878 | 26149 | 24240 |
| R-squared | 0.30 | 0.27 | 0.27 | 0.26 | 0.23 | 0.24 | 0.25 | 0.23 | 0.23 | 0.26 |

Standard errors in brackets

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

**Table 4:** Selected regression results from the extended model, 1988-2007

| | 1988 | 1989 | 1990 | 1991 | 1992 | 1993 | 1994 | 1995 | 1996 | 1997 |
|---|---|---|---|---|---|---|---|---|---|---|
| State sector employment | 0.125*** | 0.018 | 0.043 | 0.162*** | -0.070** | -0.137*** | -0.110*** | -0.081*** | -0.095*** | -0.090*** |
| | [0.0336] | [0.0343] | [0.0341] | [0.0352] | [0.0293] | [0.0291] | [0.0259] | [0.0257] | [0.0241] | [0.0216] |
| Experience 11-20 years | 0.360*** | 0.346*** | 0.360*** | 0.284*** | 0.269*** | 0.290*** | 0.272*** | 0.280*** | 0.283*** | 0.267*** |
| | [0.0109] | [0.0119] | [0.0109] | [0.0106] | [0.00951] | [0.0114] | [0.0128] | [0.0131] | [0.0136] | [0.0146] |
| Experience 21-30 years | 0.480*** | 0.467*** | 0.493*** | 0.419*** | 0.404*** | 0.440*** | 0.443*** | 0.418*** | 0.409*** | 0.395*** |
| | [0.0118] | [0.0124] | [0.0110] | [0.0108] | [0.00950] | [0.0112] | [0.0130] | [0.0132] | [0.0135] | [0.0143] |
| Experience >30 years | 0.546*** | 0.539*** | 0.577*** | 0.495*** | 0.477*** | 0.500*** | 0.511*** | 0.488*** | 0.496*** | 0.458*** |
| | [0.0130] | [0.0135] | [0.0121] | [0.0123] | [0.0110] | [0.0131] | [0.0153] | [0.0153] | [0.0158] | [0.0171] |
| College | 0.124*** | 0.146*** | 0.163*** | 0.169*** | 0.197*** | 0.241*** | 0.308*** | 0.256*** | 0.272*** | 0.261*** |
| | [0.0115] | [0.0118] | [0.0104] | [0.00998] | [0.00860] | [0.0101] | [0.0116] | [0.0116] | [0.0119] | [0.0126] |
| Technical school | 0.0659*** | 0.0818*** | 0.121*** | 0.122*** | 0.128*** | 0.151*** | 0.194*** | 0.166*** | 0.173*** | 0.157*** |
| | [0.0130] | [0.0139] | [0.0121] | [0.0116] | [0.0103] | [0.0121] | [0.0141] | [0.0143] | [0.0146] | [0.0159] |
| Senior high school | 0.0187* | 0.0157 | 0.0141 | 0.0390*** | 0.0473*** | 0.0707*** | 0.0890*** | 0.0687*** | 0.0844*** | 0.0918*** |
| | [0.0103] | [0.0108] | [0.00955] | [0.00945] | [0.00856] | [0.0101] | [0.0119] | [0.0117] | [0.0119] | [0.0126] |
| Provincial dummies | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Observations | 7729 | 6941 | 7317 | 7435 | 9648 | 9288 | 9254 | 9170 | 9220 | 9184 |
| R-squared | 0.347 | 0.33 | 0.387 | 0.377 | 0.383 | 0.373 | 0.361 | 0.335 | 0.334 | 0.317 |
| | 1998 | 1999 | 2000 | 2001 | 2002 | 2003 | 2004 | 2005 | 2006 | 2007 |
| State sector employment | -0.005 | 0.006 | -0.007 | 0.045*** | 0.143*** | 0.180*** | 0.204*** | 0.227*** | 0.205*** | 0.198*** |
| | [0.0206] | [0.0184] | [0.0174] | [0.0168] | [0.00913] | [0.00895] | [0.00814] | [0.00789] | [0.00768] | [0.00774] |
| Experience 11-20 years | 0.275*** | 0.251*** | 0.279*** | 0.272*** | 0.239*** | 0.270*** | 0.281*** | 0.238*** | 0.264*** | 0.256*** |
| | [0.0156] | [0.0162] | [0.0172] | [0.0182] | [0.0109] | [0.0112] | [0.0108] | [0.0105] | [0.0106] | [0.0110] |
| Experience 21-30 years | 0.398*** | 0.361*** | 0.387*** | 0.374*** | 0.306*** | 0.322*** | 0.305*** | 0.271*** | 0.289*** | 0.301*** |
| | [0.0152] | [0.0159] | [0.0169] | [0.0178] | [0.0107] | [0.0110] | [0.0105] | [0.0104] | [0.0105] | [0.0109] |
| Experience >30 years | 0.460*** | 0.419*** | 0.438*** | 0.402*** | 0.353*** | 0.357*** | 0.338*** | 0.278*** | 0.284*** | 0.291*** |
| | [0.0185] | [0.0182] | [0.0191] | [0.0193] | [0.0121] | [0.0122] | [0.0117] | [0.0117] | [0.0116] | [0.0117] |
| College | 0.324*** | 0.386*** | 0.449*** | 0.455*** | 0.451*** | 0.474*** | 0.469*** | 0.463*** | 0.445*** | 0.487*** |
| | [0.0134] | [0.0138] | [0.0151] | [0.0154] | [0.00958] | [0.00983] | [0.00933] | [0.00941] | [0.00943] | [0.00975] |
| Technical school | 0.189*** | 0.235*** | 0.277*** | 0.244*** | 0.233*** | 0.240*** | 0.239*** | 0.226*** | 0.226*** | 0.235*** |
| | [0.0166] | [0.0173] | [0.0196] | [0.0197] | [0.0122] | [0.0125] | [0.0120] | [0.0123] | [0.0123] | [0.0129] |
| Senior high school | 0.104*** | 0.142*** | 0.144*** | 0.177*** | 0.149*** | 0.143*** | 0.144*** | 0.150*** | 0.130*** | 0.147*** |
| | [0.0134] | [0.0139] | [0.0155] | [0.0154] | [0.00997] | [0.0102] | [0.00977] | [0.00996] | [0.00994] | [0.0102] |
| Provincial dummies | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Observations | 9002 | 8861 | 8662 | 8490 | 20926 | 23667 | 25002 | 25878 | 26149 | 24240 |
| R-squared | 0.296 | 0.274 | 0.265 | 0.257 | 0.24 | 0.255 | 0.266 | 0.255 | 0.253 | 0.275 |

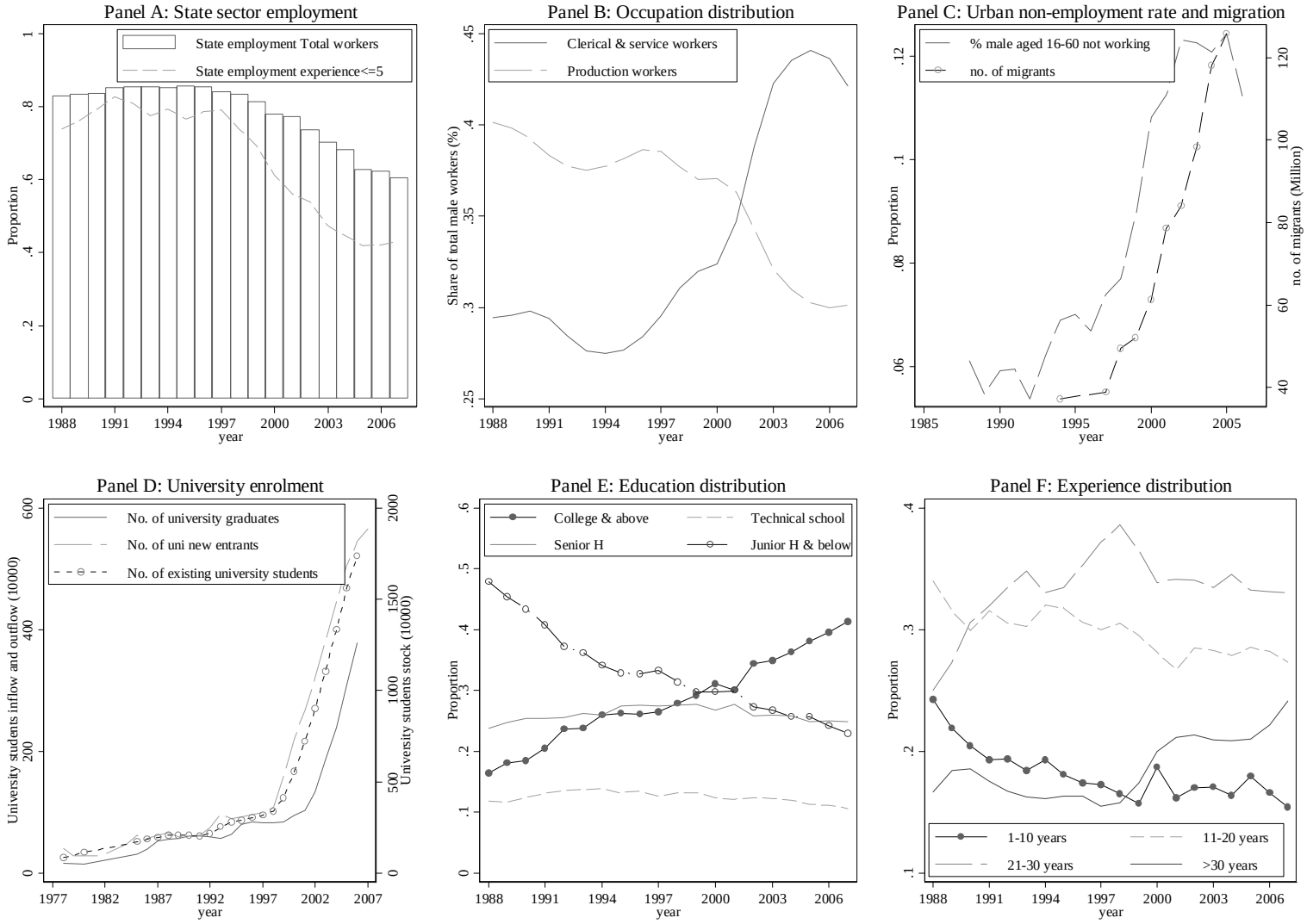
*** p<0.01, ** p<0.05, * p<0.1

Standard errors in brackets

**Table 5:** Decomposition of the variance of log annual earnings

| | 2007-1988 | | 1992-1988 | | 1998-1992 | | 2003-1998 | | 2007-2003 | |
|---|---|---|---|---|---|---|---|---|---|---|
| Basic model, 1988 as the base year | Mean | % | Mean | % | Mean | % | Mean | % | Mean | % |
| Total actual difference | 0.204 | -- | -0.026 | -- | 0.158 | -- | 0.076 | -- | -0.004 | -- |
| Observable price difference | 0.046 | 23% | 0.004 | -14% | 0.037 | 23% | 0.004 | 5% | 0.002 | -40% |
| Composition | -0.027 | -13% | -0.014 | 55% | -0.019 | -12% | -0.009 | -12% | 0.015 | -328% |
| Within variance | 0.185 | 91% | -0.015 | 60% | 0.140 | 88% | 0.081 | 107% | -0.021 | 468% |
| Basic model, 2007 as base year | | | | | | | | | | |
| Observable price difference | 0.031 | 15% | -0.005 | 20% | 0.037 | 24% | 0.002 | 3% | -0.003 | 68% |
| Composition | -0.008 | -4% | 0.002 | -7% | -0.005 | -3% | -0.007 | -10% | 0.002 | -55% |
| Within variance | 0.180 | 88% | -0.022 | 87% | 0.125 | 79% | 0.081 | 107% | -0.004 | 87% |
| Extended model, 2007 as base year | | | | | | | | | | |
| Observable price difference | 0.026 | 12% | -0.004 | 15% | 0.034 | 21% | -0.003 | -4% | -0.005 | 111% |
| Composition | 0.047 | 23% | 0.048 | -185% | 0.006 | 4% | 0.038 | 51% | 0.021 | -473% |
| Within variance | 0.136 | 65% | -0.069 | 271% | 0.118 | 75% | 0.040 | 53% | -0.021 | 462% |

**Figure 1:** Various aspects of urban labour supply and demand




Notes: Data for Panel A, B, E, and F are for 16 provinces using Urban Household Income and Expenditure Survey (UHIES) 1986-2007. For Panel C, the series for the non-labour force is from UHIES and the series for migration is from World Bank (2009). For Panel D the data are obtained from Educational Statistical Yearbook of China (Various years).

**Figure 2:** Distribution of male earnings

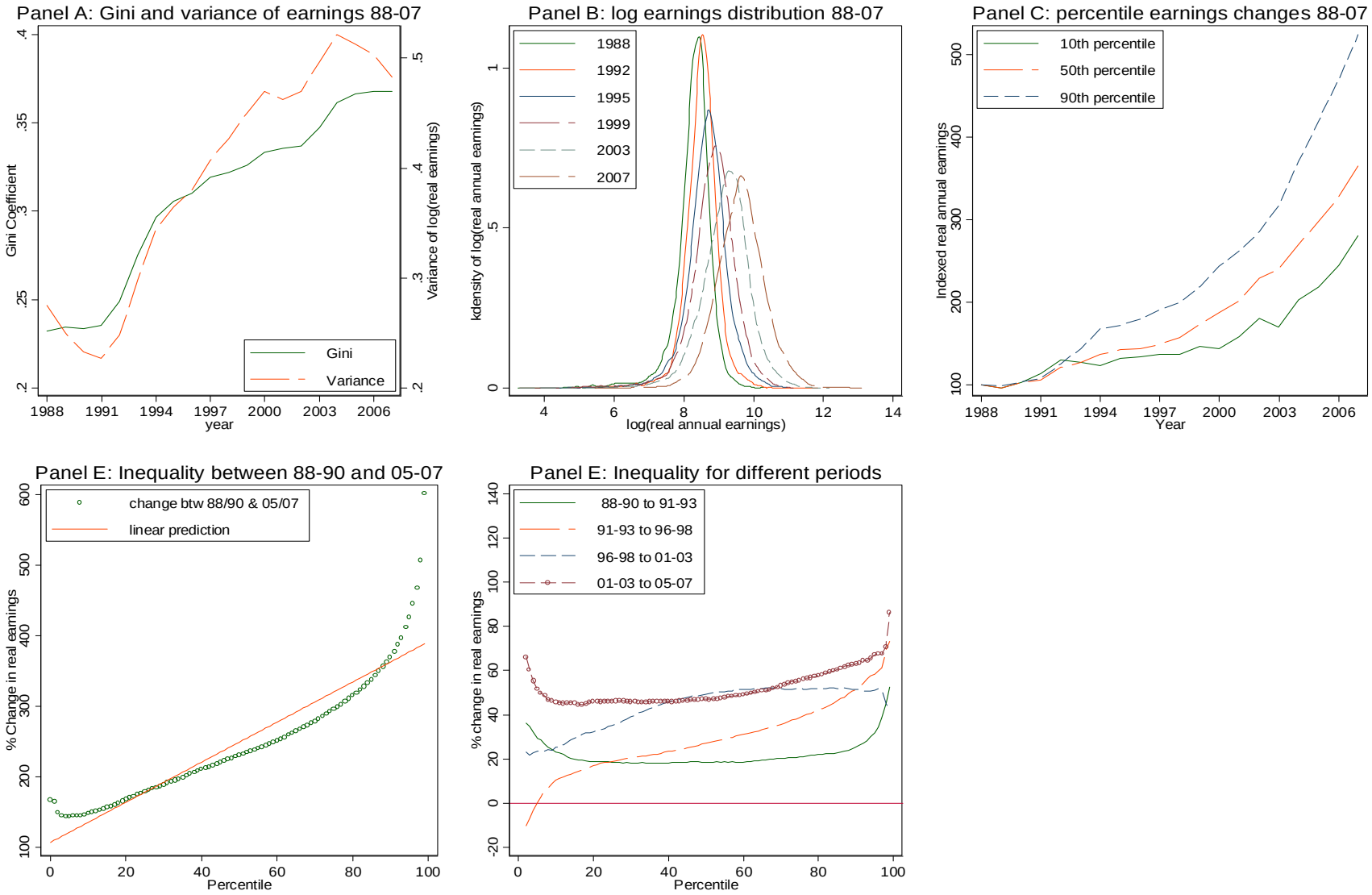


**Figure 3:** Variance of log earnings for subgroups

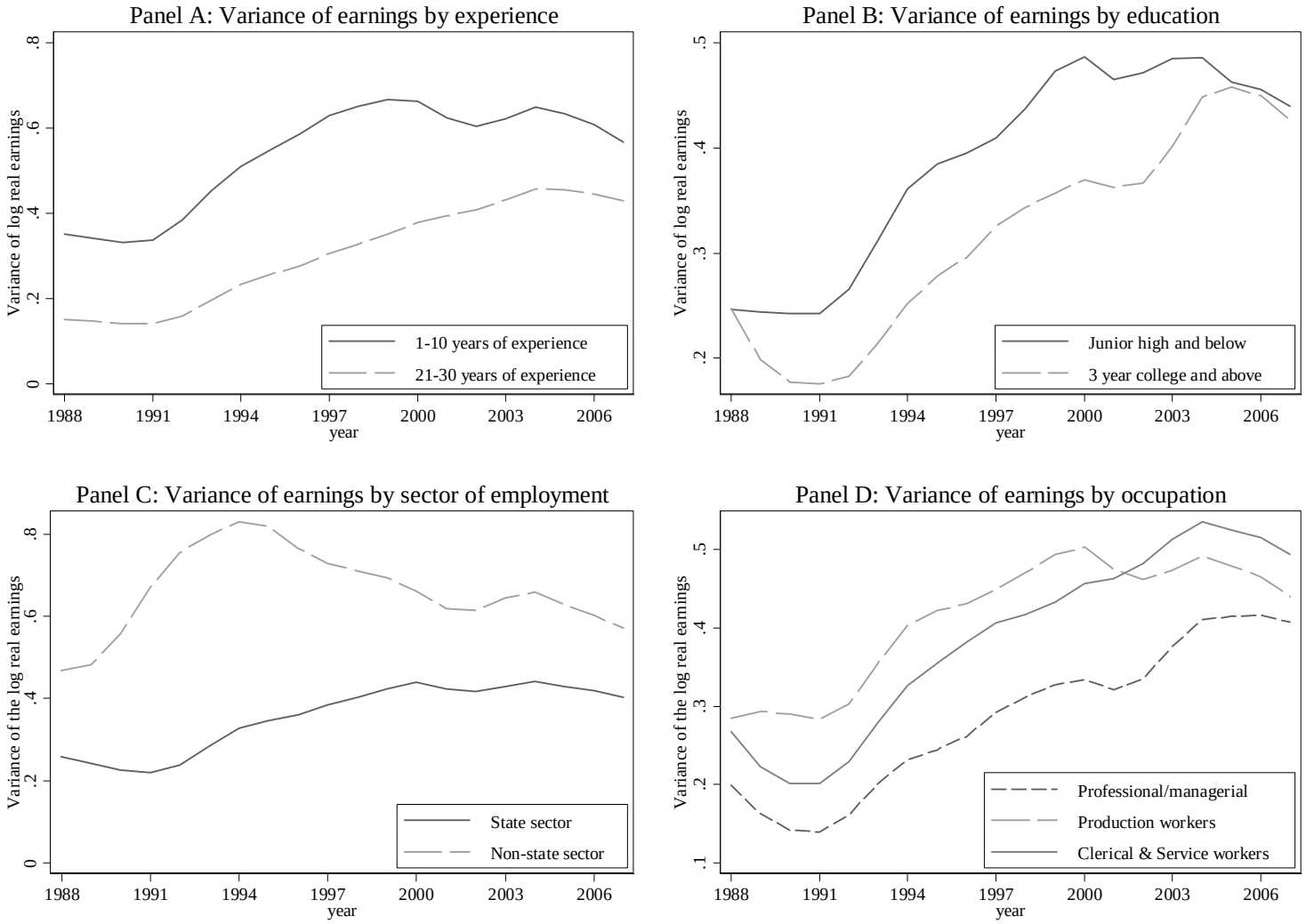


**Figure 4:** Change in returns to observable characteristics

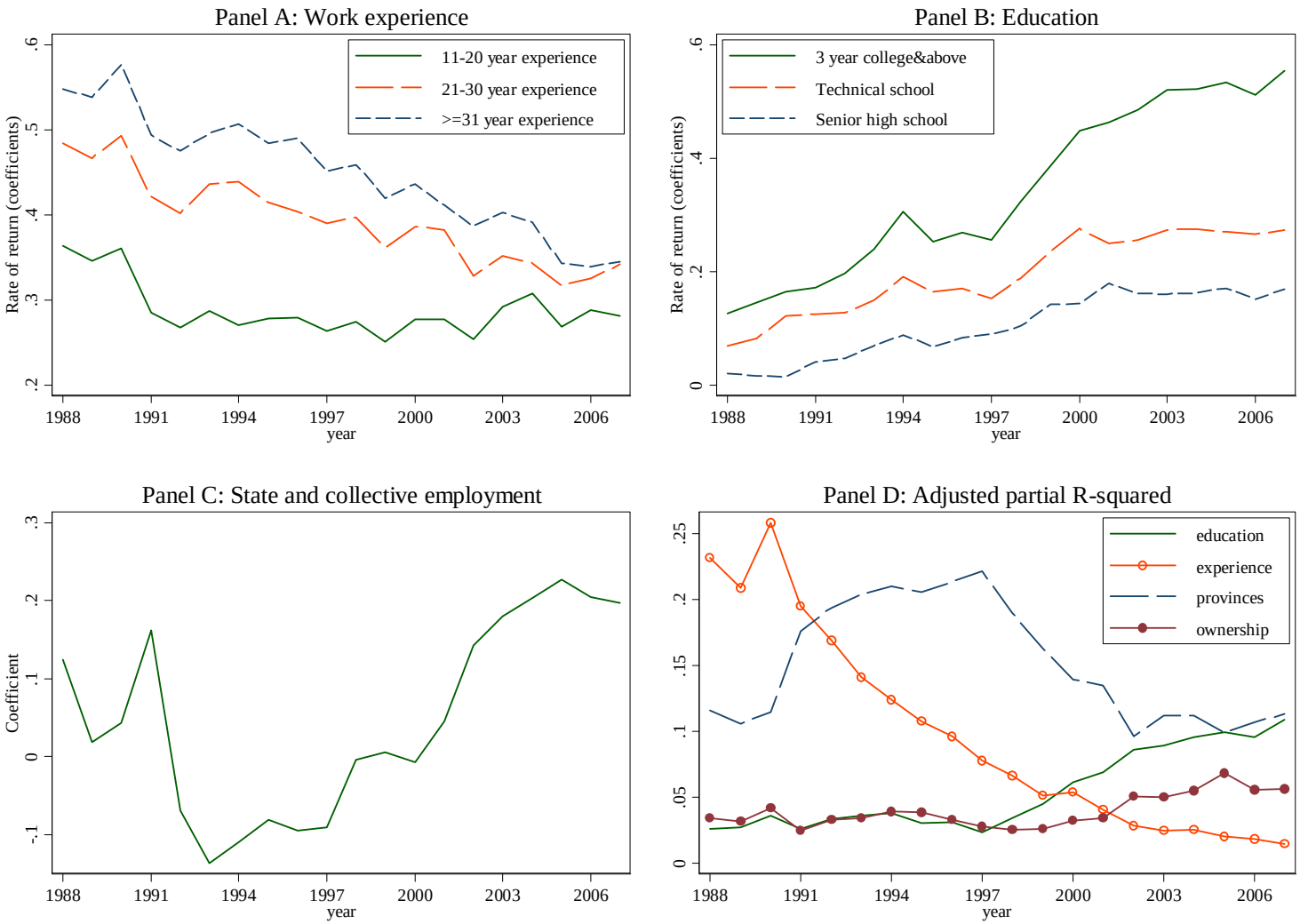


**Figure 5:** Decomposition of the change in log earnings variance

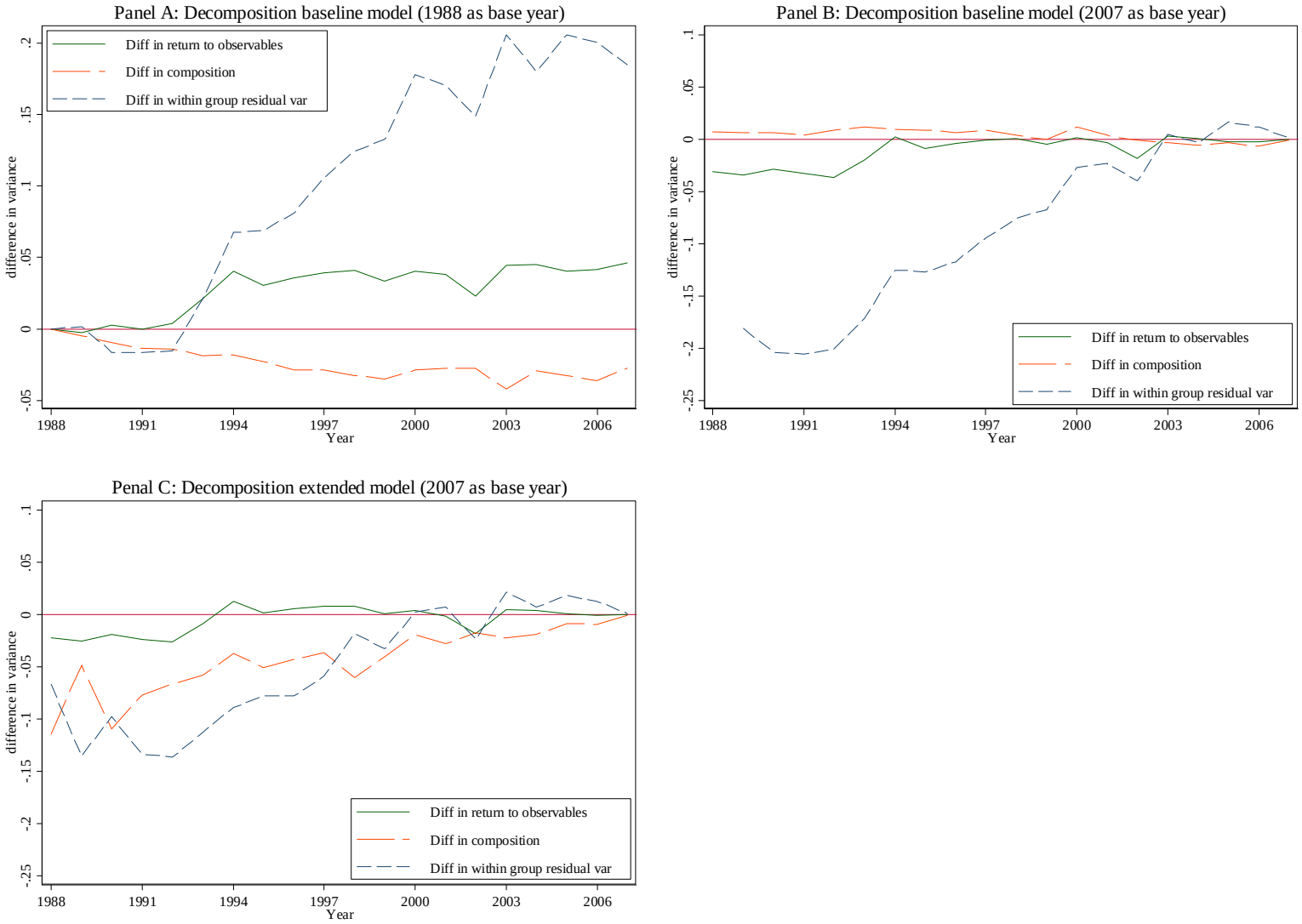


**Figure 6:** Within group residual variance Change

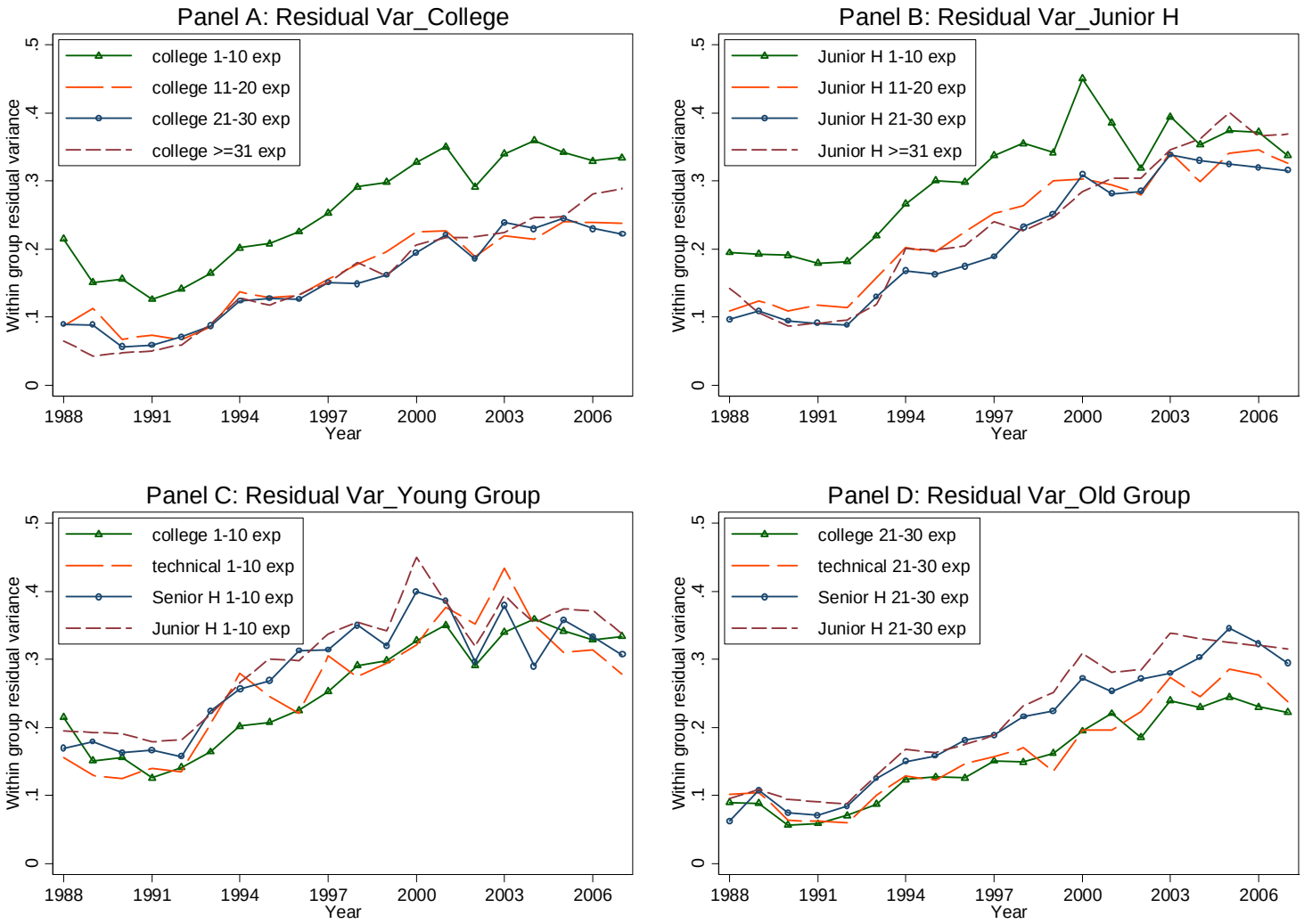


**Figure 7:** Reproduction of Panels C and D of Figure 6

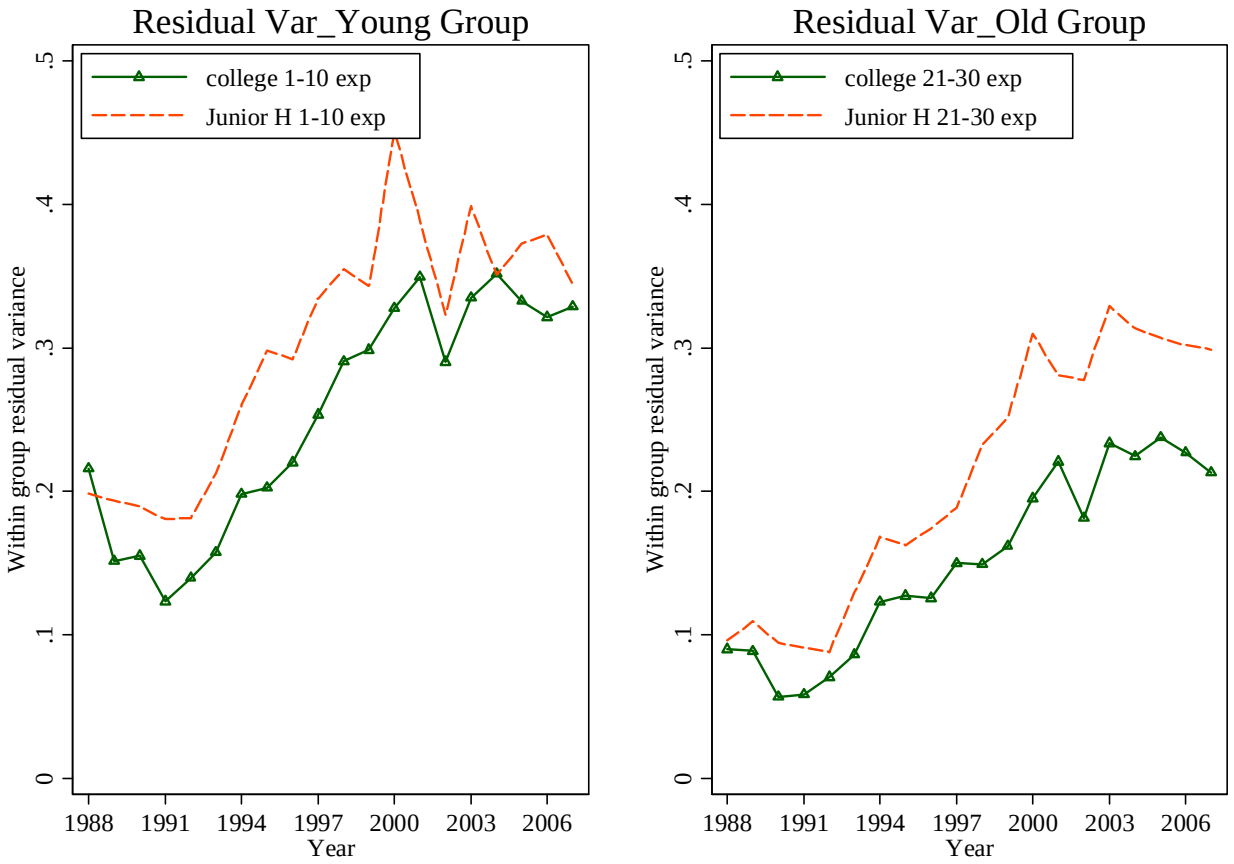


**Figure 8:** Illustration of the impact of college expansion on change in the distribution of unobserved skills

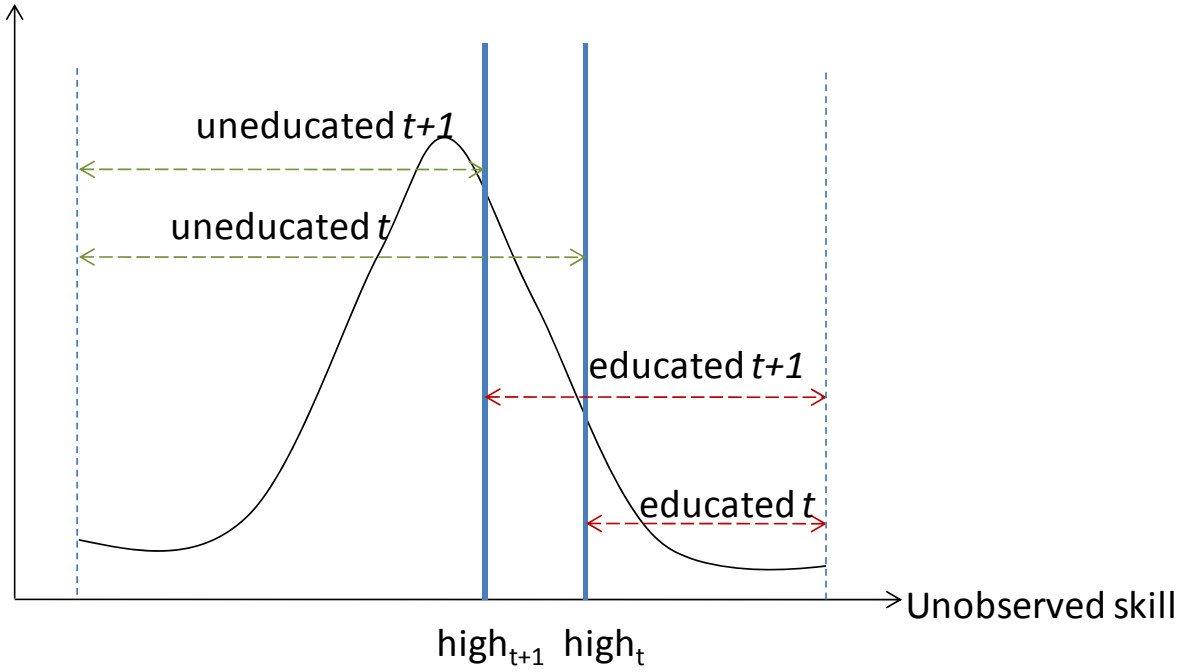
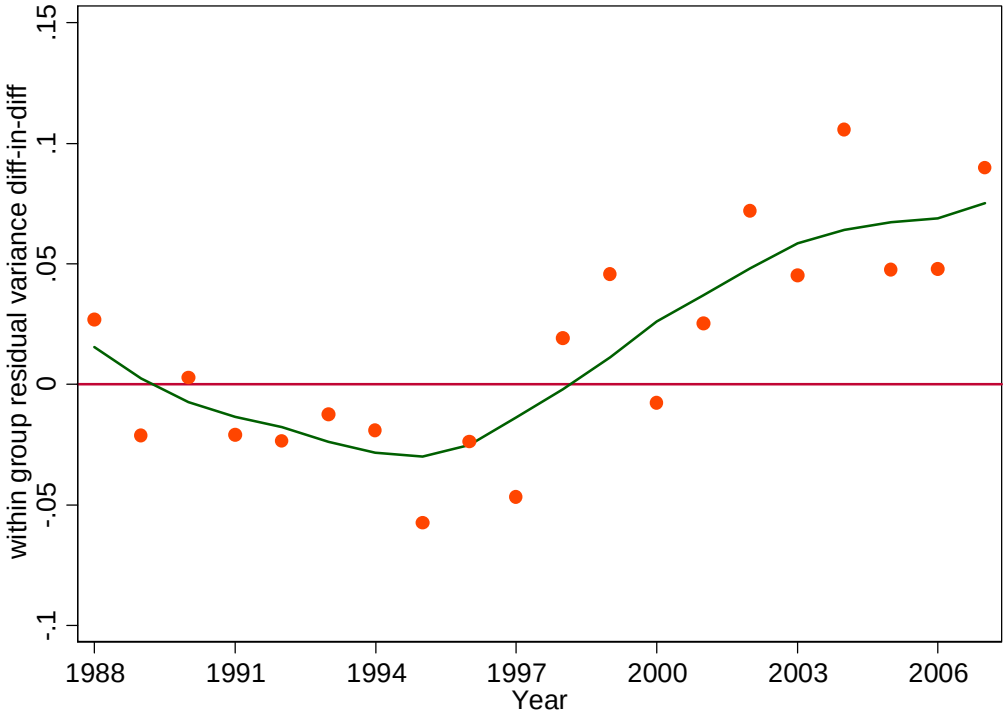


**Figure 9:** Difference-in-differences of within group residual variances



**Appendix A**: Simulation results