Residual Wage Inequality in Urban China, 1995-2007

Chunbing Xing

June 2010

Residual Wage Inequality in Urban China, 1995-2007

Chunbing Xing

Beijing Normal University
and IZA

ABSTRACT

Residual Wage Inequality in Urban China, 1995-2007

We use three waves of urban household survey data from 1995 to 2007 to investigate the trends of residual inequality and its determinants. First, we describe the change of overall and residual wage inequality over time. One important new pattern is that the rise in both the overall and residual inequality mainly happened at the upper half of the wage distributions (i.e. the rich are getting relatively much richer) from 2002 to 2007. From 1995 to 2002, however, it is truer to say that the poor are getting relatively much poorer. Second, by using two complementary semi-parametric methods, we find that the composition effect is negligible. Instead, the change in skill prices plays a dominant role in the rise of residual inequality. Finally, by constructing a panel data at the city level, we find that ownership restructuring is an important factor that has caused the skill price to rise, especially in the earlier period. Another finding is that China's export share of GDP has a positive effect on the enlargement of the upper half distributions. This effect is more significant in the latter period from 2002 to 2007, highlighting the impact of China's entry into the WTO.

JEL Classification: J3

Keywords: urban China, residual inequality, decomposition

Corresponding author:

Chunbing Xing
Beijing Normal University
in 2010 on leave at:
University of Western Ontario
Room 4012, Social Science Centre
London, Ontario, N6A 5C2
Canada
E-mail: xingchunbing@gmail.com

Residual Wage Inequality in Urban China, 1995-2007

1. Introduction

China's urban wage inequality increased rapidly in the last two decades (Chen and Ravallion, 2007; Park et. al, 2003). According to the estimates of Li, et. al, (2007), the Gini coefficients of urban wages increased from 0.238 in 1988 to 0.364 in 2003, other measures of inequality also increased significantly. Not surprisingly, the increase in wage inequality attracts much attention (see for example Li and Sato, 2006, Riskin, Zhao, and Li, 2001).

Most of the existing research on China focuses on wage gaps *between* well defined groups. Although the *between inequality* is important, there are still large proportions of wage variation that can't be explained by those observable characteristics. We define the latter part as *residual inequality* or *within inequality*. In terms of a standard Mincerian wage equation, where wage (in log form) is modeled as a function of education, experience, plus an error term, *between inequality* can be captured by the distribution of these observable characteristics and their coefficients, and residual inequality is captured by the error term. It's well known that the explanatory power of variables such as education and experience is seldom more than 30 percent.[1] Meanwhile, both casual observations and academic research indicate that within group difference is becoming more and more manifested both in China (Cai et al, 2010) and other countries. Therefore, understanding how the unexplained (residual) part evolves and exploring the underlying reasons are important for understanding the overall inequality.

The increase in residual inequality may be due to several reasons. The first and also the most important reason is the increase in the price of unobservable skills. Holding the skill distribution constant, rise in skill price will increase residual inequality. In fact, the view that the increase in unobserved skill price caused the increase in residual inequality is a dominant one in U.S. (Juhn, et. al, 1993, Acemoglu, 2002, Katz and Autor, 1999). However, when looking more into details of the structure of residual inequality, we may come up with very different conclusions. Notice that the overall residual distribution (inequality) is simply the results of conditional residual distribution integrated over observable characteristics. Loosely speaking, the overall residual inequality is a weighted average of residual inequalities of various well defined groups, and the weight is the share of these groups in the whole sample. Also notice that within inequalities of different groups are not necessarily identical. Therefore, composition change in observable characteristics alone can cause change in the overall residual inequality. Taking this into consideration, Lemieux (2006) finds that the increase in residual inequality in US is partly due to composition effect. Besides price effect and composition effect,

---

[1] In Cai et al (2010)'s research, the explanatory power of observables in income regressions is around 33%. They include sector, occupation, ownership and provincial dummies in their regressions, as well as age, age squared, and education.

other factors that may have impact on residual inequality include measurement error, and change of unobservable skill distribution (Meng, et al, 2010)[2]. What's the underlying forces driving the increase of residual inequality is still hotly debated.

In this paper, we use three waves of urban household survey data from 1995 to 2007, to investigate the trends of residual inequality and its determinants. First, we describe overall and residual wage inequality in detail. From 2002 to 2007, one important new pattern is that, the rise in both the overall and residual inequality mainly happened at the upper half of the wage distributions (i.e. the rich are getting much richer). In the former period from 1995 to 2002, however, it's truer o say that the poor are getting relatively much poorer. This phenomenon has seldom been studied.

Second, by constructing counterfactual residual wage inequality using two semi parametric methods, namely DFL (DiNardo, et al, 1996) and quantile regression based approach (Machado and Mata, 2005[3], Angrist, et al, 2004), we explicitly consider the effect of composition change on residual inequality. Surprisingly, although skill composition changed significantly, both approaches indicate that composition effect is negligible. This result provides empirical support for those researches that directly link the change in residual inequality with change in skill price, without considering composition effect. We would like to point out the beginning that the price effect may include the effect of changes in measurement error and unobservable skill distributions. As these two effects are more difficult to address, we don't make effort to separate them out. However, some additional exercises in the latter part of this paper indicate these two effects may not be very large.[4]

Finally, by construction a panel data at the city level, we find that ownership restructuring is an important factor that causing the skill price to rise, especially in the earlier period. This explanation is fundamentally different from that for western countries, which emphasizes skill biased technological change. At the same time, this result is consistent with most of the existing research on China emphasizing the role of institutional change (Zhang et al, 2005, Meng et al, 2010, Cai et al, 2010). Another finding is that China's entry into WTO has a positive effect on the enlargement of upper half distributions. This finding helps to differentiate this paper from the existing ones further. Most of the existing research either simply declined to consider globalization as a potential explanatory variable, or dismiss it as unimportant (Zhang et al, 2005 for example). At this moment, I am only aware of one research that emphasizing globalization's effect on urban inequality, Cai et al, (2010). However, they only use variance as inequality measure, and haven't seen the effect of globalization on different part of the residual distribution. By pooling all data from 1992 to 2003 together, they haven't documented the structural change accompanying China's entry into WTO neither.

---

[2] To be more rigorous, there are two types of composition effects. One is due to the change in observable characteristics distribution, the other is due to change in distributions of unobservable skills. The first one is our focus in this paper.
[3] We use the terms "MM approach" and "quantile regression based approach" interchangeably.
[4] This can also be thought of as making assumptions that measurement error and unobservable skill distributions haven't change significantly. Given the major transition in terms of education expansion, these assumptions are strong. However, we haven't found significant evidence against these assumptions.

Although the research on residual inequality of urban China is relatively few, we've already seen several notable exceptions. The most recent related researches are Cai et al, (2010) and Meng et al (2010). Using the data from the Urban Household Income and Expenditure Survey (UHIES), Cai et al (2010) decompose the income inequality into between-group inequality and within-group inequality. They find within-group inequality accounts for over 60% of the overall inequality. They also uncover explanations (economic restructuring, urbanization and globalization) for the acceleration of urban inequality. As their focus is on urban income and consumption inequality, they don't consider the price effect versus composition effect issue. Meng et al (2010)'s data comes from the same source. But their data covers only 16 provinces while that of Cai et al (2010) covers all the provinces. Instead of using a broad income concept as in Cai et al, (2010), they consider only earnings income[5]. By decomposing the increase in earnings overall variance into observable price effect, composition effect and within cell variance, they find that within group inequality is the major force that driving up the overall earning inequality. However, they don't consider the composition effect and price effect of residual inequality explicitly either. Another common shortcoming of these two studies is that their decomposition exercises are based on variance decomposition. While variance is an important measure of inequality, it can't give a complete description if the residual distributions are not normal. By using two complementary semi-parametric approaches, we can consider the whole distribution of residual distributions.

Two relatively earlier researches are Park et. al, (2003), and Li, et. al, (2007). Park et. al, (2003) analyzed the increase of urban wage inequality from 1988 to 1999. They find that there was a substantial increase in within-group inequality which is robust to numerous group definitions. The returns to unobservable skills as measured by regression residuals explain much of the increase in overall inequality. Instead of using directly the methodology proposed by Juhn, Pierce, and Murphy (1993) as Park, et. al (2003), Li, et. al, (2007) used the extended JMP methods, namely quantile JMP decomposition.[6] Using the same dataset, there's no surprise that they got the same results as Park, et. al, (2003). Compared to these researches, our dataset is more recent, which allows to detect some new patterns. Another common shortcoming of these researches is their lack of explicit explanation for the rise in skill prices, which is exactly what we've done in this paper.

This paper is organized as follows. Section 2 briefly summarizes some backgrounds that may have effect on (residual) wage distribution and describes our data. Section 3 introduces methodologically how we decompose the residual wage inequality. In section 4, we present our decomposition results, emphasizing the important role of skill price change in causing the rise of residual inequality. In section 5, we aggregate our data at the city level and use fixed effect model to see what factors have effect on the change of residual inequality (and also skill prices). Section 6 concludes.

2. Background and Data Description

---

[5] The income data Cai et al, (2010) uses include wage earnings, business income, asset income and transfer income, Cai et al, (2010, p389).
[6] The quantile JMP decomposition method is proposed by Autor, Katz, and Kearney, (2005).

2.1 Economic Reform in the Late 1990s and in the New Century

From 1995 to 2007, China witnessed the deepening of reform in the urban areas. There are several wonderful works that analyze this process and its impact on the labor market and income distributions[7]. It's impossible to give a detailed and comprehensive description of this period. In the following, we will summarize three aspects that might have influence on China's urban labor market and therefore inequality.

The first, and maybe the most important one, is the radical ownership restructuring of urban sector commenced in 1997. As early as 1992, the central government has setup the goal of "establishing socialist market economy". The reform was stagnant however, with lots of SOEs having redundant employment and extremely low profits (or even loss making) in the mid 1990s (Knight and Song, 2008, p224). The major ownership restructuring happened after 1997. On the 15th National Congress of Communist Party of China, it was reiterated that private sector is an important part of our socialist market economy, and that private property right should be protected by law. At the same time, the concept of "public ownership" which is essential for *socialist* economy was reinterpreted. The point is that "public ownership" can be realized through various ways such as joint ownership, and that China can still be a socialist economy if the government controlled large scaled enterprises in key sectors (Jiang, 1997). As a consequence, lots of medium and small sized SOEs were privatized through various ways including merger, acquisition, and joint ownership. One major part of this restructuring is large amount of laid-off or unemployed workers. Most of them were absorbed by the private sector subsequently.

For at least three reasons, the restructuring in the late 1990s would have significant effects on the wage structure and therefore on residual wage inequality. Firstly, it is well established that the public sectors and private sectors have different wage determination mechanisms, with the private sector being more market oriented (see Meng, 2000, Dong and Bowles, 2002, Xing, 2008). As more workers are absorbed by the private sector, the wage structure will change consequently. The second reason is that the wage determination mechanism within the public sector also evolved, especially after the ownership restructuring. For example, Xing (2007) finds that the returns to education in public sectors increased significantly in the late 1990s as compared to earlier period. Entering into 21 century, the evolution of wage determination mechanisms within the state owned sectors continued. Beginning from around 2003, most SOEs performed another round of compensation reform, aiming to make compensations more sensitive to work performance. Thirdly and related to the second one, the restructuring in the late 1990s created quite a few monopolistic SOEs. Without competitors in the market, most SOEs profit much higher. In terms of employment, entering the SOEs is a very selective process, and most of the entering labor forces go to the private sector. There's no severe redundancy problem as compared to the 1990s. As a result, the wage gap between SOEs and private sectors increased a lot.[8] We'll investigate how the ownership restructuring affects the residual

---

[7] Comprehensive analysis are summarized in Li and Sato (2006), Riskin, et al, (2001) and Gustafsson, et al, (2009).
[8] Another reform was initiated by State-owned Assets Supervision and Administration Commission of the State Council

inequality in an indirect way in section 5.

Another major event at the turn of the century is China's accession into WTO, which paced up China's globalization process.[9] As a member of WTO, China cut down tariff rate for over 5000 products in 2002. During the same period, there was substantial trade deregulation, narrowing down of scope of quota limits, rising inflow of FDI (Wan et al, 2007). Also, China experienced tremendous surge in export following entrance of WTO. Although there have already been efforts to relate China's general globalization process with inequality,[10] little has been done to see the effect of China's entry into WTO.[11] By looking into the whole distribution, and considering structural change accompanying WTO accession, we provide more evidence on this relationship.[12]

The third major event that might affect wage inequality is China's higher education expansion. As documented in Li and Xing (2010), the radical higher education expansion policy commenced in 1999 changed dramatically the composition of urban labor force. If different education groups have different within-group wage inequalities, the expansion will influence the overall residual inequality. This is what we call "composition effect" in the next section. However, there may be another type of composition effect that is more subtle but also important. The college entrance selection system in China selects those with higher ability into college. With the expansion, candidates with relatively lower ability can be admitted into college. Without the expansion, they will stay in the high school graduate group. Within the high school graduate group, they are in the upper part of the ability or skill distribution. All these mean that education expansion will change the distribution of abilities or unobserved skills. Under the assumption that unobserved skill distribution is unchanged, the real change will be counted as price effect, therefore bias our decomposition results. Fortunately, we don't find significant change in ability distributions.

No doubt, it's impossible to exhaust all the forces that may have effect on wage inequality during this period. We ignore them for several reasons. The first and most important reason is data unavailability. Secondly, if these ignored factors are not correlated with both dependent and independent variables, ignoring them will not bias our estimates of the relationships between other factors and inequalities. Third, some seemingly important phenomenon may have little impact on our results. For example, China's urban labor market experienced huge inflow of rural migrants. However, as the labor market is segmented the migrants may have little impact on urban residents' wages (Knight and Yueh, 2004, Liu and Zhao, 2009). As our focus here is wage inequality for those with urban hukou, i.e. those officially registered in urban

---

(SASACSC) in 2004. This reform aims at regulate how the CEOs or managers are compensated.
[9] China also witness deepening of globalization in the 1990s. Wan et al, (2007) gave a brief description on *China's journey to globalization* in terms of trade and tourism, FDI, movement of people. See also Brenstetter and Lardy (2006).
[10] For example, Zhang and Zhang (2003) find that globalization is an important factor contributing to the widening regional inequality. However, Wei and Wu (2003) find that greater openness caused declination of rural urban income gap. Another paper by Wan et al (2007) also finds that globalization has a decreasing effect on regional inequality. Generally speaking, these papers focused on the relationship between globalization and regional disparity, and they used aggregated data.
[11] It's widely recognized that globalization will have major impact on inequality (Hanson, 2004 on Mexico, Topalova, 2005 on India, Attanasio et al, 2003 on Columbia).
[12] More closely related to ours, Cai et al (2010) indicate that both FDI and export have positive effects on within variance as well as on overall variance.

areas, ignoring this may not affect our results much. Finally, as will be clear in section 5, to study the underlying reasons for residual inequality change, we'll use a fixed effect model at city level. This will help us a lot by controlling for many time invariant factors.

2.2 Data and the Trends of (Residual) Wage Inequality in Urban China

In this paper, we use data from the three waves China Household Income Project (CHIP) survey in 1995, 2002, and 2007. CHIP data is well known for its high quality and national representativeness. In 1995, 2002, and 2007, the urban survey covered 11, 12, and 16 provinces, including a wide variety of regions in terms of geography and economic development. Although it is not a panel, the coverage of the survey didn't change much. In 1995 and 2002, the survey collected information in the same provinces. The only difference is that Chongqing is separated out from Sichuan Province. The survey in 2007 included more provinces while retaining those in the previous surveys. Not only most of the provinces are identical, but also the cities. Therefore, we can construct a panel data at city level when uncovering the reason of rising residual inequality, which allows us to control time-invariant unobservables.

We keep only those with wage data greater than zero, including both male and female. To focus more on the labor market, we use labor income of urban workers, including wages, subsidies and labor incomes from other sources. All the income data is deflated into 1995 RMB using the national CPI. We first consider the changing patterns of the wage levels and wage inequalities. Table 1 reports various measures of wage inequality as well as means. For male, average wages increased by 0.49 and 0.52 log points respectively from 1995 to 2002 and from 2002 to 2007. The increase in the latter period is higher. The increases for female are relatively lower than male especially in the latter period (0.48 and 0.36 log points respectively).

Next, we focus on inequality measures. For male, the variance of log wages increased from 0.41 in 1995 to 0.54 in 2002, and to 0.78 in 2007.[13] The differentials of log wages between 50th and 10th percentiles increased from 0.72 to 0.87, while those between 90th and 50th percentiles increased from 0.60 to 0.68. Over 65 percent of the increase in the overall inequality as measured by the differential between 90th and 10th percentiles happened in the lower half. This pattern reversed completely in the second period. Around 60% of the increase in the 90-10 percentile differential happened in the upper half. In words, the inequality increased mainly because the poor were left behind in the first period, in the second however, it is mainly because the rich got richer much faster. For female, inequality also increased in both periods. Unlike their male counterparts, the increase in inequality concentrated in the upper half in the first phase (1995-2002), but in the lower half in the second phase (2002-2007).

The increase in inequality may be due to many reasons. Just to see the importance of residuals, we use the JMP approach to decompose the change of overall wage inequality into three parts, namely differences in observable quantities, differences in observable prices, and differences in

---

[13] This is inconsistent with Meng et al, (2010). They find that variance inequality stopped rising in recent years. This may due to the fact that we include all labor incomes. If we don't include the labor income other than wages and subsidies, the variance in 2007 would be much lower. Other inequality measures are not very sensitive to this part of income however.

unobservable quantities and prices.[14] The results are reported in table 2. It's clear that price effect can explain the bulk of the increase at specific percentiles for both male and female. This finding is consistent with other researches on China, for example, Li, et. al, (2007) and Park et al, (2006). When we come to the change of differential between different percentiles, differences in unobservable quantities and prices play a very important role. For example, from 1995 to 2002, it can explain 60-70 percent of the increase of the wage differentials for male.

To see the importance of residual further, we predict residuals after running wage regressions for male and female by years. In terms of variance, the observables (education, experience, their full interactions, and region dummies) can at most explain one quarter of the total variance for both gender and for various years, with the explaining power varying a little bit at different positions of the distribution and in different years (the residual distribution is given by Figure 1). This claim is still true even after we controlled ownership, sector, and occupation dummies. Therefore, it's not unexpected that after purging out the between-group differentials, the wage dispersions are still large. For male, the residual inequality patterns are very similar to the overall wage inequality. For female, the pattern is more similar to their male counterpart when considering residuals. In particular, the increase of residual inequality from 2002 to 2007 concentrated on the upper half of the residual distribution, which is in contrast with the change of overall wage inequality during the same period.

Residual inequality will be closely related the observables if its conditional distribution is dependent on these variables. To see this more clearly, we divide the sample of each year into 24 education-experience cells for both male and female. We have 6 categories for education, namely primary and below, junior middle school, high school, technical school, professional school, college and above, and 4 categories for experience, namely below 10 years, 11-20 years, 21-30 years, 31 years and above. The so-called conditional distribution is in essence the within group variations.

Within the narrowly defined groups (education/experience cells), the variance of wages almost unanimously increased during the first period for males. The increasing trend remained for majority of these subgroups. The exceptions are subgroups of lower education and lower experience levels. Subgroups with education level lower than technical school (inclusive) and with less than 10 years of experience witness the largest decrease in within group inequality. At a given point of time, the within group variance vary with education and experience subgroups. In all three years, variances of wages are relatively larger for subgroups with lower education and lower experience levels. The above patterns for male are also applicable to female.

Given these descriptive results, three points are worth mentioning regarding the within group inequalities. First, the patterns we've got here are at odds with both theoretical predictions and vast majority of empirical evidences in developed countries (see Lemieux (2006) for example). We believe this is mainly due to the fact that the wage determination mechanism in China is

---

[14] JMP approach has shortcomings. It has been criticized for not taking into consideration of heteroskedasticity which is extremely important for the discussion of residual inequality. Further, it doesn't consider explicitly the effect of composition change on residual inequality. It fact, this effect is included in the contribution due to differences in observable quantities.

different from developed market economy. Under the traditional regime, significant share of labor is in public sector, where wages are set according to rigid rules, which is heavily dependent on education and experience (seniority). Note that the more educated and more experienced are more likely to be employed in public sector, it's natural that their within group dispersion is lower. With the deepening of China's economic reform, employers have more autonomy to set wages according to the employees' ability other than education and experience. Therefore, we can expect the dispersions of higher skills would increase more.

Second, the fact that wage dispersions of different groups are different means that changes in distribution of observable skills are potentially able to have effect on changes in residual inequality. This is exactly the so-called "composition effect". From table 3 and table 4, it's clear that the labor force became more educated and more experienced during first period. During the second period however, the trend is not as obvious. Another way to see this is to run regressions with year dummies as dependent variables. We pool the data of 2002 with that of 1995 and 2007 respectively. Let the dichotomous dependent variable equal to one if the individual belongs to the sample of 2002 and zero otherwise. The results reported in table 5 confirmed our conclusion that the skill composition did change significantly. The results also indicate that most of the change occurred in the first period, and the change for female are more significant.

Finally, we only calculate within group variance at this moment. However, variance is only one measure of conditional distribution and it may be sensitive to extreme values. A more systematic way to describe the conditional distributions and their evolutions is to estimate series of quantile regressions. This point will be much clearer in the methodological section. Therefore, we postpone the discussion on quantile regressions and corresponding results to latter sections.

3. Decomposing Residual Wage Inequality

We consider log form of individual wages. Furthermore, we assume that wages are determined according to a Mincer type wage equation:

(1) $y_{it} = X_{it}\beta_t + \varepsilon_{it}$

Where y_{it} is the log form of real wages, X_{it} stands for observable skills such as categories of potential experience (age-years of schooling-6), education levels, their interaction terms, and region dummies. ε_{it} is the error term. According to JMP, the residual is a product of some unobserved skills, e_{it}, and the return to unobserved skills, p_t. Taking into consideration of measurement error v_{it}, Chay and Lee (2000), Lemieux (2006) considered the following error component model:

(2) $\varepsilon_{it} = p_t e_{it} + v_{it}$

Residual wage inequality is simply the measured inequality in the residual ε_{it}. While may researches only focus on variance of ε_{it},[15] we consider the functional of residual distributions, which can be thought of as more general measures of residual inequality. In fact, the

---

[15] Lemieux (2006), Meng et al (2010), Cai et al (2010) for example.

reweighting approach or the *quantile regression based approach* (see below) for accounting for the composition effect are applicable in very general settings. Now, consider the marginal distribution of the wage residual.

(3) $F_t(\varepsilon_i) = \int F_{\varepsilon|X}(\varepsilon_i \mid X, t_\varepsilon = t) dG_X(X \mid t_X = t)$

The unconditional distribution of wage residual can be written as the conditional distribution of wage residuals given observed skills X at time t integrated over the whole distribution of X at time t. Equation (3) highlight the fact that residual distribution (inequality) may be different for different observed skills. Before moving forward, it's worthwhile to discuss a little bit more about the dependence of wage dispersion (residual) on X.

When we use variance as the inequality measure, the dependence of residual distribution on X means heteroskedasticity (See Lemieux, 2006b for more detailed discussion). The wide use of quantile regressions to estimate wage equations nowadays reflects researchers' concern about the existence of it (Koenker and Bassett, 1982, Buchinsky, 1994, 1998, AKK, 2005). And in fact, various theories predict that the dispersion of wages should increase with education and potential experience. For example, Mincer (1974) pointed out differential investments in on-the-job training will cause the wage dispersion increase as a function of experience. Farber and Gibbons (1996) reach the same conclusion in a simple learning model. And using a random coefficient model, and in the spirit of Becker (1967) and Mincer (1997), Lemieux (2006a) shows that the variance of wages should be larger for more- than less-educated workers, and that the variance should increase more for more- than less-educated workers when the price of education increases. The evidence from China also suggests heteroskedasiticity (Xing, 2007, Luo, 2008). It's worth noting however, that the pattern of wage dispersion within different education/experience cells is quite different from that of U.S. We'll come back on this later.

Next, assume that measurement error v_{it} and unobserved skill e_{it} do not change over time.[16] Equation (3) can be rewritten as:

(3') $F_t(\varepsilon_i) = \int F_{\varepsilon|X}(p_t e_{it} + v_{it} \mid X, t_p = t) dG_X(X \mid t_X = t)$

The increase in residual inequality will be solely due to rising skill price p_t and the composition change of skills, i.e. the change in G(X|t). One central task of this paper is to determine to what extent does the increase in residual inequality can be attributable to rising prices or composition change.

To account for the effect of composition change on residual distribution, it's essential to construct a counterfactual distribution that keeps skill composition or skill price constant through the period under consideration. Consider the change of residual inequality from t=0 to t=1. What would the residual distribution would be like in time 1 if the skill composition is constant at time 0? The counterfactual distribution can be written as follows: [17]

---

[16] This is a strong assumption. We will see to what extent does this assumption holds in section 5. For convenience, we pretend this assumption hold as Lemieux (2006b).

[17] We need to assume that the change of skill distributions from $G_X(X \mid t_X = 0)$ to $G_X(X \mid t_X = 1)$ have no effect on the conditional distribution of residuals, so that the counterfactual distribution can be written as equation (4).

$$(4) \quad F_C(\varepsilon_i) = F_c\left(\varepsilon_i \,|\, t_X = 0, t_p = 1\right)$$

$$= \int F_{\varepsilon|X}\left(p_t e_{it} + v_{it} \,|\, X, t_p = 1\right) dG_X\left(X \,|\, t_X = 0\right)$$

Note that the counterfactual distribution is simply the conditional distribution (or skill price) of time 1 integrated over the skill distribution of time 0. Based on this observation, two approaches are proposed to construct such counterfactuals.

The first approach is based on quantile regression and re-sampling. By running quantile regressions for different quantiles of time 1, we can have a detailed description of the conditional distribution. The estimated skill price structure (conditional distribution) can be applied to skill distribution of time 0. This can be done using a re-sampling method proposed by Machado and Mata (2005), or simply by multiplying the quantile coefficient matrix with the data matrix as proposed by Angrist et al, (2004).[18] As is clear in the estimation process, this approach can estimate the conditional distribution *explicitly*, with some risk of putting strong restrictions on the structure of conditional distributions. Meanwhile, by using re-sampling method or its equivalent, no effort is needed to estimate the composition change parametrically.

An alternative approach originally proposed by Dinardo, Fortin, and Lemiuex (1996) is just like a mirror image of the above one. The insight of the DFL approach is that counterfactual distribution can be estimated by reweighting the residual sample of time 1 using the following weight (see Dinardo, et. al, (1996) for details),

$$(5) \quad \theta = \Pr(t = 0 \,|\, X) / \left(1 - \Pr(t = 0 \,|\, X)\right)$$

Pooling the data of time 0 and time 1 together, $\Pr(t=0|X)$ can be estimated using the logit or probit models. Using the predicted probability, the weight can be calculated for each observation (predicted residual) in time 1, $\hat{\theta}_i = \widehat{\Pr}(t = 0 \,|\, X_i) / \left(1 - \widehat{\Pr}(t = 0 \,|\, X_i)\right)$. It can then be used to estimate the counterfactual density and to calculate various counterfactual inequality measures.

It's clear that the parametric model and its predictions are used to estimate the composition change *explicitly*, and the skill price of time 1 is just hiding in the residual data of time 1. To be more intuitive, consider the skill compositions of year 1995 and 2002. With the gradual expansion of higher education, we can expect more college graduates in the 2002 sample than in 1995 (Li, et al, 2008). Using the residual wage sample of 2002 to construct the counterfactual holding skill composition at the 1995 level, we should give college graduates less weight. Note that the probability of a graduates being in the 1995 sample will be relatively lower than that in the 2002 sample holding other characteristics constant. The weight of an observation with college degree will be less than one.

These two approaches, DFL and MM, are conceptually similar. For example, Autor, Katz, and

---

[18] To simulate the actual distribution and construct counterfactual distributions, Angrist,et al, (2004) show that greater precision can be obtained by multiplying the entire g(x) distribution by the quantile regression coefficient matrix. Therefore, we use the latter approach instead of the re-sampling approach proposed by Machado and Mata (2005). Autor et al, (2005) also use the former approach. To be concise, we sometime call the quantile regression based approach MM approach.

Kearny (2005) extended the MM approach to investigate composition effect and price effect in the change of wage inequality. To compare their results to Lemieux (2005), who used reweighting approach, they find its substantive differences with the latter are *not* consequential for their conclusions (They do draw different conclusions, but not because of different methods). As neither one can claim global superiority over the other, we use them both. We would like to point out at this earlier stage that the results from these two approaches are quite similar. As the MM approach gives more volatile results and is more time consuming, after presenting the benchmark results, we rely more on the DFL reweighting approach to do decomposition exercises.

Finally, we define a real-valued functional v(F). v(.) can be thought of as rules to map different distributions (Fs) to different real numbers. For example, the v(.) can be a rule to calculate the Gini coefficients of different distributions. It can also be a rule to calculate many other statistical measures of F, such as variance, Theil indices, percentiles, the differential between 90th and 10th percentiles, etc. Having constructed the counterfactual distribution of residuals, and letting $v_0 = v\left(F_0\left(\varepsilon_i\right)\right)$, $v_1 = v\left(F_1\left(\varepsilon_i\right)\right)$ and $v_C = v\left(F_C\left(\varepsilon_i\right)\right)$, it's straight forward to decompose the change of any distribution statistics $v\left(F\right)$ into two parts:

(6) $v_1 - v_0 = \left[v_1 - v_c\right] + \left[v_c - v_0\right] = \Delta_X^v + \Delta_P^v$

The term in the first bracket is the change due to composition effect (Δ_X), while that in the second bracket is the change due to skill price effect (Δ_P).

4. Counterfactual Residual Inequality and Decomposition Results

4.1 Conditional Distribution and Quantile Regression Based Decomposition

We've already seen how within group variance varies among different groups in section 2. To see how the conditional distribution of residual wages depends on observable characteristics in a more systematic way, we estimate quantile regressions for both genders and for each year at various conditional quantiles including (1, 1.5, 2, …, 98, 98.5, 99). If the conditional distributions of wage residuals don't depend on observable characteristics, there shouldn't be systematic difference in coefficients among different percentiles. The results from quantile regressions indicate that this is not the case. In particular, we estimate returns to various levels of education and experience.[19] The results for male and for different years are reported in Figure 2a to Figure 2c. In Figure 2a, one obvious feature is that the coefficients in lower quantile regressions are higher than those in upper quantile regressions. Therefore, the coefficient-quantile profiles are downward sloping for all the variables. This pattern changed significantly in 2002. For lower levels of educations, the relationship between coefficients in different quanitle regressions and quantiles become inverse U shaped. For higher levels of education (professional college, and college graduates), the relationships are still monotonic and downward sloping, but to a lesser extent. The same is

---

[19] For ease of interpretation, we do not consider the interaction terms between education and experience here. But in constructing the counterfactual using quantile regressions, we do include interaction terms.

also true for experience dummies. In 2007, the coefficient-quantile relationships become even more flatter. The results for female show similar patterns, and to save space, they are not reported.

Again, the evolution of quantile regression results indicate that the conditional distribution (skill price) changed dramatically from 1995 to 2007. Meanwhile, the systematic differentials between different quantiles confirm that the conditional residual distributions at various levels of observables are not necessarily identical, and therefore composition change may also have substantative effect on the overall residual inequality.

Next, we construct conterfactual residual distributions to evaluate the relative importance of price effect and composition effect. Take the counterfactual with 2002 skill price and 1995 composition for example. First, we estimate quantile regressions using the 2002 data. In all the quantile regressions, residual wage is the dependent variable and education dummies, experience category dummies, their interactions, and region dummies are regressors. The quantile regressions are run at the (1, 1.5, 2, …, 98, 98.5, 99)th percentiles. After estimation of each quantile regression, the coefficients are applied to the 1995 data (using out-of-sample prediction). As we run quantile regressions 197 times, the 1995 data are used 197 times to predict residual wages. All these predicted residuals are then used to estimate the couterfactual residual distributions. The main results are reported in Figure 3. In figure 3a, the blue line is the counterfactual distribution with the combination of 2002 skill price (quantile regression coefficients) and 1995 composition. Similarly, the green dash line is the counterfactual combining 2002 skill price and 2007 skill composition. These two lines almost overlap with the factual distribution for year 2002. Notice that all these three lines are based on skill price of year 2002, and the only differences are due to composition change. The results indicate that composition effects are very small. For female, we have the same conclusion.

4.2 DFL Decomposition

Next, we also apply the reweighting approach to estimate factual and counterfactual residual kernel densities. The predicted probabilities from the above probit models are used to calculate the weight according to equation (5). We have two counterfactuals and one factual residual distribution. The two counterfactuals are constructed using the residual data of 2002. By reweighting the observations according to composition change from 1995 to 2002, and from 2002 to 2007, we are holding the price of unobservable skills constant at the 2002 level and letting the skill composition vary.

It turns out that the composition effect can explain little of the change in residual inequality. Figure 4 indicates that the counterfactual distributions for 1995 and 2007 are very close to the factual one in 2002. This means that the composition change have little effect on the change of residual inequality. This is true for both genders.

An alternative way to see the role of price effect and composition effect is to see the change of percentiles. Take male between 1995 and 2002 for example. In Figure 5a, we plot the differential between 1995 and 2002 different percentiles. The solid line is the

differential of factual percentiles, which can be thought of as overall change. The long dash line depicts the differential between the factual percentiles of 2002 and the counterfactual percentiles of 1995, which is due to composition change. The dash-dot line depicts the differential between the factual and counterfactual percentiles of 1995, which is due to price change. The latter one is very close to the factual differential, indicating the predominant role of price change for enlarging residual inequality. This composition effect is positive at lower percentiles and negative at higher percentiles. Thus composition change tends to *decrease* residual inequality. Although the composition effect plays a minor role on the whole, it seems to have much larger effect at the very low end of the distribution. This is also true for female during the period between 1995 and 2002. In the second period, the changing pattern of the overall inequality is different from the first period, which we already discussed at the data description section. The role of price effect and composition effect didn't change much however, with composition change playing an even more minor role.

To see the robustness of our results, we calculate counterfactual inequality measures using different combination of skill prices and skill compositions. In table 6, each column is under the price of a specific year, and each row is under the skill composition of a specific year. Take the element at the first column and second row for example. The inequality measure is calculated using the counterfactual density with skill price of 1995 and composition pattern of 2002. And so on so forth. Therefore, the change along the diagonal is the factual change, while the change of inequality along columns is solely due to composition change and the change along rows is due to price effect. Obviously, price change almost always explains off the overall change. Composition change plays a minor role, if not negligible.

5. Caveat and Explanations

5.1 Change of Unobserved Skills or Measurement Error?

In section 2, we make two assumptions. The first is that the distribution of unobserved skills *within* specific groups doesn't change overtime. The second is that measurement error doesn't change either. Whether these assumptions hold or to what extent they hold have strong implications for the interpretation of our results. It's clear that we can't separate out the price effect if the distribution of unobserved skill (e_{it}) or (and) that of the measurement error (v_{it}) change overtime.

With education expansion, it's natural to worry that unobserved skill distributions for well defined education-experience groups change overtime. As mentioned in the background section, there was a significant post-secondary education expansion in 1999 and thereafter. As the number of new students admitted to colleges or universities increased, we expect that the unobserved skill distributions within both high school graduates and college or university graduates will change accordingly, and this will contaminate the price effect.

Surprisingly, we haven't found that within college inequality increased due to the change in unobserved skill dispersion, neither have we found skill distribution change for high school graduates. In figure 3, we calculate inequality measures (variance, the difference between the 90th and 50th percentiles, and that between the 50th and 10th percentiles) using a 1/5 random sample from 1% census data in 2005. Given that much more high school graduates are admitted to college or universities, we expect there will be much more heterogeneity within those groups, and therefore the within group inequalities will increase. However both for male (Figure 6a) and female (Figure 6b), we haven't found obvious increase in inequalities for the younger groups. Although the composition change is substantial (see figure 1 in Li and Xing, 2010), there are no obvious discontinuities in the inequalities. In fact, the inequality measures are lower for those with few market experiences. For high school graduates, within group inequality is lower for the younger groups, but there're no discontinuities as we expect. As Figure 6 uses only one cross sectional data set, we can't separate the expansion effect from the age or cohort effect. According to human capital theory, the within group inequality for younger workers is lower. Therefore the younger age effect may counter balance the expansion effect. However, if this is true, the change in unobservable skill distributions will have little effect on our decomposition also. The reason is that even in 2007, those groups that affected by the expansion policy are still young. The change in unobservable skill distribution needs time to have effect.

In Figure 7, we perform similar exercises using the three CHIP surveys. As the samples are small, we put the male and female sample together. But still, the figures have more volatility, especially for the college samples. Although being volatile, the inequality-experience profiles for year 2002 and for 2007 are quite similar, especially for the younger groups. For those older groups, within group inequalities in 2007 are larger than that in 2002 and than that in 1995. Yet for those younger groups, it's hard to say that college or university within group inequalities have increased because of expansion. It's also hard to say that within group inequalities for younger high school graduate has declined. Therefore, evidences both from the census data and from CHIP data show that the change in unobservable skill distributions won't be a major problem that will contaminate our decomposition.

As for measurement errors, it's hard to judge whether it becomes more manifested or not. However, the fact that these three surveys were conducted by the same research team alleviates this concern. Furthermore, there seems no significant change in the *volatility* of those inequality-experience profiles in Figure 7.

5.2 Potential Explanations

The basic point of our strategy to uncover the underlying reasons for the rise in residual inequality is to explore the regional variations. To do that, we first run wage regressions for different years as in section 3 and 4. Using the predicted residuals, we calculate the residual inequality measures at the city level. These residual inequality measures are then used as dependent variables to be explained. According to the background section, we consider three independent variables. To reflect the effect of ownership restructuring, the *soe* variable is the

fraction of people in the state owned sector. A *college share* variable is constructed to reflect the proportion of people with college or university degrees. And finally, to reflect the possible effect of globalization on inequality, we calculate the proportion of export in regional GDP. For reason of data availability, the third variable is at provincial level, while the first two are constructed at the city level. A nice feature of the CHIP data is that many cities participated in all the three surveys. Therefore, we can construct a panel dataset at the city level and estimate fixed effect models. To keep the panel balanced, we drop those cities that participated in the survey less than three times. There are 33 cities that appear in all the surveys.

The results are reported in Table 7. Panel A is the results for males. As expected, most of the coefficients for *soe* are significantly negative. As more and more labor are allocated into the private sector, the residual inequality increased. According to our decomposition results, the rise in residual inequality is mainly due to price effect. The results here indicate that the ownership restructuring increased skill prices. This result is consistent with existing literature that studying the evolutions of return to educations. For example, Zhang et al, (2005) find that returns increased from 1988 to 2001, Li and Ding (2003) also find this pattern. The fact that we are looking into residual inequality differentiate our research from theirs.

By considering the two periods (from 1995 to 2002, and from 2002 to 2007) separately, we find that ownership restructuring has more significant effect during the first period. This is consistent with the fact that the most radical change happened in the late 1990s. In addition, by considering the upper half inequality and lower half inequality separately, we find that the ownership restructuring has more effect on the lower half.

As emphasized in section 2, a relative new phenomenon of urban wage structure is that majority of the residual inequality happened at the upper half. One important finding in Table 7 is that the trade variable (*exp/gdp)*) has significant effects on the upper half of the residual distributions during the second period. Therefore the results here not only confirm our hypothesis that China's entry into WTO has effect on urban (residual) wage inequality, but also explain (at least partly) why the inequality on the upper half rises more rapidly from 2002 to 2007. Therefore, our results here shed new lights on how globalization affects inequality from a very different dimension.

In most of the cases, coefficients for variable *college share* are not significant. This reinforced our conclusion that the higher education expansion has little effect on our estimation and interpretation of the decomposition results. Most of the general pattern for males is also true for female. One major exception is that exp/gdp variable has significant effect on the change of the lower half of residual distributions (panel B of table 7). The effect is positive but not significant for the upper halves. Another exception is that SOE variable has more significant effect for the upper half instead of the lower ones.

In panel C and D, we try another way to construct inequalities at the city level. Instead of run wage regression for the whole sample for particular years, we run wage regressions for each city and for different years. City level residuals are predicted and residual inequality measures

are constructed after that. Ownership restructuring and exports are still important factors that driving up residual inequalities. And the patterns in C and D are similar to those in panel A and B. Other variables are not significant any more, therefore are not reported. Finally, we also do exercises without requiring a balanced panel data. The sample increases a lot. There's no essential difference between the balanced panel data results and the unbalanced ones, however. This makes us more confident about our results.

5.3 Labor Market Segmentation and Residual Inequality

Existing researches have shown that China's labor market is segmented (Knight and Song, 2008 for example). This will have a strong implication for the interpretation of price effect in residual inequality. Given the existence of labor market segmentation, individuals with identical characteristics both observable and unobservable, may have significantly different earnings simply due to the fact that they are in different occupations, industries, or ownerships. Therefore, the price effect we've established above includes both *real price* of unobservable skill and *rents* between different sectors. However, it's more realistic to say that individuals' sector affiliation is often closely related to education levels and unobservable ability. Therefore, it's notoriously difficult to separate out the *rent* effect from the *real price* effect.

The exercise we've done here is to further include the industry, occupation, and ownership dummies when predicting the residuals.[20] As already been shown, residual inequalities are still large even after further controlling these variables (see Table 1). In the decomposition results for male reported in Table 6, price change plays a major, if not dominant, role. Composition change plays a more important, but still minor, role. The composition effect is larger in the period from 2002 to 2007. The above conclusions are also true for female, but with some irregularities. The results in part B of Table 6 highlight the fact that different sectors have different wage dispersions, and that there have been significant restructuring in terms of ownership, industry, and occupations. Fully accounting for this requires more efforts and spaces, we leave this for future research.

6. Conclusion

The time from 1995 to 2007 is a very important period for China's economic reform and development. The urban sector changed dramatically. In terms of wage income, not only the differential between different well defined population increased, but also that within groups are increased. Theoretically, this increase in residual inequality however can't be simply regarded as rise in skill prices. However, by using the DFL reweighting and quantile regression based approaches, we confirm that the rise in urban residual inequality is mainly due to price effect.

---

[20] To what extent the segmentation is correlated with unobservable ability also determines how far we should go when defining *residuals* (In cases where segmentation is perfectly correlated with observable characteristics such as education and experience, the discussion here will be trivial.) Two extreme cases will help to illustrate this point. If segmentation is perfectly correlated with unobservable skills, the seemingly segmentation is just a result of different individuals selecting or being selected by different sectors. In this case, segmentation is just a superficial phenomenon and will not be a concern at all. Another extreme case is that segmentation is totally uncorrelated with unobservable skills, and the differences between different sectors are just *rents*. In this case, the industry, occupations, and ownerships dummies should be included in the wage equation when predicating residuals.

The explanations for the rise in skill price are fundamentally different from those for western countries, however. The later emphasize skill biased technological change. The reasons that we have found for China are mainly due to China's transitional feature. We emphasize two. One is ownership restructuring. The other is China's integration into the world market. Regarding the first, our results are consistent with most of the existing research that studying prices of observable characteristics. The second add new evidence on the relationship between China's openness and urban wage structure, by emphasizing the structural change brought by China's entry into WTO, and by emphasizing its bigger impact on upper half of residual distributions.

Considering the major expansion higher education, the results with little composition effect really surprised us. What also surprised us is that there's no obvious evidence showing that the unobserved skill distributions within particular groups changed after the expansion. My conjecture is that these effects need time to show up. Those who are affected by the expansion are still young in our samples. According to human capital theories, as people gaining more experience, their income level will diverge. To see the full effect of the expansion needs more time and new data.

References:

Acemoglu, Daron, .Technical Change, Inequality, and the Labor Market,. *Journal of Economic Literature* 40, March 2002, 7-72

Angrist, Joshua, David Autor and Victor Chernozhukov. 2004 "Quantile Wage Decompositions: Estimation and Inference, with an Application to the U.S. Minimum Wage." Work in Progress, Massachusetts Institute of Technology.

Autor, David, Lawrence Katz, and M. S. Kearney. (2005), "Trends in U.S. Wage Inequality: Re-Assessing the Revisionists." NBER Working Paper No. 11627.

Becker, Gary S. *Woytinski Lecture*, University of Michigan, 1967. Reproduced in *Human capital (2nd Edition),* University of Chicago Press, 1975.

Bishop, J., Luo, F., and Wang, F., 2005, "Economic Transition, Gender Bias, and Distribution of Earnings in China," *Economics of Transition*, vol. 13, pp:239-259.

Buchinsky, Moshe,1994, "Changes in the U.S. Wage Structure 1963-1987: Application of Quantile Regression", *Econometrica*, Vol. 62, No. 2., Mar., pp:405-458.

Buchinsky, Moshe, 1998, "The Dynamics of Changes in the Female Wage Distribution in the USA: A Quantile Regression Approach," *Journal of Applied Econometrics*, Vol. 13, No. 1. (Jan. - Feb., 1998), pp: 1-30.

Cai, Hongbin, Yuyu Chen, and Li-an Zhou, 2010, "Income and Consumption Inequality in Urban China: 1992-2003", *Economic Development and Cultural Change*, 58:385–413.

Card, David, 2009, "Immigration and Inequality", CReAM Discussion Paper Series, CDP No 07/09.

Dong, Xiaoyuan and P. Bowles, (2002), "Segmentation and Discrimination in China's Emerging Industrial Labor Market." *China Economic Review*, Vol.13, 170-196.

DiNardo, J., N. Fortin, T. Lemieux, 1996, "Labor market institutions and the distribution of

wages, 1973～1992: A semi-parametric approach", *Econometrica,* 64 (5): 1001～1044.

Farber, H. and R. Gibbons, "Learning and Wage Dynamics," *Quarterly Journal of Economics* 111, 1996, 1007-47

Gustafsson Björn,Shi Li,Terry Sicular, (eds), 2008, *Inequality and Public Policy in China*, Cambridge University Press.

Hanson Gordon H, 2004, "Globalization, Labor Income, and Poverty in Mexico", NBER working paper.

Jiang, Zemin, 1997, "Holding up the Flag of Deng Xiaoping's Theory, Opening a New Period of Socialist Economy in the New Century (*Gaoju deng Xiaoping lilun weida qizhi, ba jianshe you zhongguo tese shehuizhuyi shiye quanmian tuixiang xinshiji*)", http://cpc.people.com.cn/GB/64162/64168/64568/65445/4526285.html.

Juhn, Chinhui, Kevin M. Murphy, and Brooks Pierce (1993) .Wage Inequality and the Rise in Returns to Skill,. *Journal of Political Economy* 101, June 1993, 410-42

Katz, Lawrence, and David Autor, 1999, "Changes in the Wage Structure and Earnings Inequality", in O. Ashenfelter and D. Card (eds.) *Handbook of Labor Economics*, volume 3A, 1999, Amsterdam: Elsevier Science

Knight, John and Song, Lina, 2003, "Increasing Urban Wage Inequality in China: Extent,

Elements and Evaluation," *Economics of Transition,* vol. 11, no. 4, pp: 597-619

Knight John, and Song Lina, 2008, "China's Emerging Urban Wage Structure, 1995-2002", in Gustafsson Björn,Shi Li,Terry Sicular, (eds), 2008, *Inequality and Public Policy in China*, Cambridge University Press.

Knight John and L. Yueh, 2004, "Job Mobility of Residents and Migrants in Urban China", *Journal of Comparative Economics*, 32(4), 637-60.

Koenker, Roger and Gilbert Bassett, Jr., 1978, "Regression Quantiles," *Econometrica*, Vol. 46, No. 3, pp: 33-50.

Koenker, Roger and Gilbert Bassett, Jr., 1982, "Robust Tests for Heteroscedasticity Based on Regression Quantiles," *Econometrica*, Vol. 50, No. 1, pp:43-62.

Li, Shi, and Hiroshi Sato, (eds.) 2006, Unemployment, Inequality and Poverty in Urban China, Routledge.

Li Shi and Ding Sai, 2003, "Long-term Change in Private Returns to Education in Urban China", *Social Science in China*, No. 6. (Chinese)

Li Shi and Xing Chunbing, 2010, "China's Great Leap Forward in Higher Education and its Labor Market Consequences", Working Paper.

Li, Yao, John Whalley, Shunming Zhang, and Xiliang Zhao, 200?, "The Higher Educational Transformation of China and Its Global Implications", NBER Working Paper.

Li, Xiaohua, Yaohui Zhao, and Lili Lu, 2007, "Effects of Education on Wage Inequality in Urban China, 1988-2003", *The 6th PEP Research Network General Meeting*.

Liu Xuejun and Zhao Yaohui, 2009, "The Impact of Labor Migration on Urban Labor Markets in China", *China Economic Quarterly*, No.2. (Chinese)

Lemieux, Thomas, (2006a), "Postsecondary Education and Increasing Wage Inequality", *the American Economic Review*, Vol(96), No. 2.

Lemieux, Thomas, (2006b), "Increasing Residual Wage Inequality: Composition Effects, Noisy Data, or Rising Demand for Skill?" *the American Economic Review*, Vol(96), No. 3. June.

Lemieux, Thomas. 2008. "The Changing Nature of Wage Inequality." *Journal of Population Economics*, 21(1): 21-48.

Luo Chuliang, The return to education and its distribution in urban China, *China Economic Journal*,2008, vol. 1(2).

Meng, Xin. (2000). *Labor Market Reform in China*, Cambridge University Press

Meng, Xin, Kailing Shen, and Xue Sen, 2010, "Economic Reform, Education Expansion, and Earnings Inequality for Urban Males in China, 1988-2007", *IZA Discussion Paper No. 4919.*

Mincer, Jacob, *Schooling, Experience, and Earnings*, New York: NBER, 1974

Mincer, Jacob, 1997, "Changes in Wage Inequality, 1970-1990", *Research in Labor Economics 16*, 1997, 1-18.

Orazio Attanasio Pinelopi K. Goldberg Nina Pavcnik, 2003, "Trade Reforms and Wage Inequality in Colombia", working paper.

Park, Albert, Xiaoqing Song, Junsen Zhang, Yaohui Zhao, 2006, "Returns to Skill, Labor Market Transition, and the Rise of Wage Inequality in Urban China, 1988 to 2003", Working Paper.

Petia Topalova, 2005, "Trade Liberalization, Poverty and Inequality: Evidence From Indian

Districts", NBER Working Paper #11614.

Riskin C., Zhao R. and Li S., 2001, (eds.) *China's Retreat from Equality: Income Distribution and Economic Transition*, Armonk, N.Y., M.E. Sharpe, 2001

Wei Shangjin and Wu Yi, 2003, "Globalization and Inequality without difference in data definition, legal system and other institutions", NBER working paper.

Wan Guanghua, Lu Ming and Chen Zhao, "Globalization and Regional Income Inequality in China, Empirical Evidence from within China", *Review of Income and Wealth*, 53(3), 35-59

Xing Chunbing, 2008, "Human Capital and Wage Determination in Different Ownerships, 1989–97", in Wan Guanghua (ed.), *Understanding Inequality and Poverty in China: Methods and Applications*, Palgrave Macmillan, Feb., 2008.

Xing Chunbing, 2007, "Wage Determination and Returns to Education in Different Ownerships of China: Evidence from Quantile Regressions", *Frontiers of Economics in China*, Volume 2, Number 1, March, 2007, 114-136.

Yin, Xiangshuo, 1998, *The Procedure and Effects of China's Reform of International Trade*, (in Chinese), Shanxi: Shanxi Economic Publishing House.

Zhang Xiaobo, Zhang Kevin H., 2003. "How does globalization affect regional inequality within a developing country? Evidence from China", *Journal of Development Studies* 39(4): 47-67.

Zhang, Junsen, Zhao, Yaohui, Albert Park, and Song Xiaoqing, (2005), "Economic returns to schooling in urban China, 1988 to 2001," *Journal of Comparative Economics*, 33, 730-752.

,

Table 1 Wage inequality and residual inequality

	1995	2002	2007	1995	2002	2007
	Male			Female		
real log wage						
Mean	8.63	9.12	9.64	8.40	8.88	9.24
90-10	1.31	1.55	1.88	1.52	1.69	2.05
50-10	0.72	0.87	1.00	0.89	0.91	1.14
90-50	0.60	0.68	0.88	0.63	0.78	0.92
Variance	0.41	0.54	0.78	0.61	0.59	1.05
residual of log wage (Residual I)						
90-10	1.14	1.37	1.58	1.32	1.44	1.80
50-10	0.60	0.77	0.86	0.75	0.81	1.00
90-50	0.54	0.60	0.73	0.57	0.63	0.81
variance	0.33	0.42	0.61	0.51	0.46	0.83
Residual with more controls (Residual II)						
90-10	1.11	1.29	1.45	1.23	1.38	1.50
50-10	0.57	0.70	0.75	0.69	0.76	0.78
90-50	0.53	0.59	0.70	0.54	0.62	0.71
variance	0.31	0.38	0.46	0.42	0.40	0.47

Table 2 Juhn-Murphy-Pierce decomposition

	Male					Female			
1995-2002	T	Q	P	U		T	Q	P	U
mean	0.490	0.049	0.442	-0.001		0.482	0.090	0.393	-0.001
p10	0.374	0.099	0.363	-0.088		0.424	0.142	0.343	-0.061
p50	0.528	0.046	0.458	0.023		0.439	0.066	0.392	-0.018
p90	0.614	0.043	0.494	0.077		0.593	0.091	0.464	0.037
d9010	0.240	-0.056	0.130	0.165		0.169	-0.051	0.122	0.098
d5010	0.154	-0.053	0.095	0.112		0.015	-0.076	0.049	0.043
d9050	0.086	-0.003	0.035	0.053		0.154	0.025	0.073	0.056
2002-2007									
mean	0.521	0.089	0.432	-0.001		0.356	0.096	0.260	-0.001
p10	0.368	0.044	0.399	-0.075		0.197	0.069	0.258	-0.131
p50	0.500	0.065	0.424	0.010		0.422	0.106	0.263	0.054
p90	0.692	0.120	0.478	0.094		0.555	0.118	0.285	0.152
d9010	0.324	0.076	0.079	0.168		0.358	0.048	0.027	0.282
d5010	0.132	0.021	0.025	0.085		0.225	0.036	0.004	0.185
d9050	0.192	0.055	0.054	0.083		0.133	0.012	0.023	0.098

Note:

T = Total difference

Q = Contribution of differences in observable quantities

P = Contribution of differences in observable prices

U = Contribution of differences in unobservable quantities and prices

Table 3 Within-group variance and composition change, male

	Within-group variance			change			Labor force share			change	
	1995	**2002**	**2007**	**95-02**	**02-07**		**1995**	**2002**	**2007**	**95-02**	**02-07**
EDU=1											
0-10	0.76	1.87	0.32	1.11	-1.56		0.14	0.19	0.15	0.04	-0.04
11-20	0.35	0.44	0.81	0.09	0.37		0.53	0.28	0.22	-0.25	-0.06
21-30	0.57	0.43	0.59	-0.14	0.16		1.86	0.78	0.55	-1.08	-0.24
31+	0.83	0.44	1.03	-0.40	0.59		1.91	1.29	1.23	-0.62	-0.06
EDU=2											
0-10	0.78	0.90	0.52	0.13	-0.39		3.78	1.68	1.48	-2.09	-0.20
11-20	0.52	0.86	0.83	0.33	-0.03		6.55	4.93	3.10	-1.61	-1.83
21-30	0.27	0.47	0.65	0.20	0.18		12.73	8.71	6.30	-4.02	-2.40
31+	0.61	0.49	1.17	-0.12	0.68		4.91	8.63	10.05	3.73	1.42
EDU=3											
0-10	0.48	0.69	0.57	0.21	-0.12		5.29	3.92	2.65	-1.36	-1.27
11-20	0.33	0.57	0.61	0.24	0.04		9.22	6.73	4.95	-2.49	-1.77
21-30	0.29	0.38	0.81	0.09	0.44		5.59	12.20	10.11	6.61	-2.09
31+	0.44	0.44	0.80	0.00	0.36		2.45	3.63	7.00	1.17	3.37
EDU=4											
0-10	0.38	0.69	0.52	0.31	-0.17		3.82	2.19	2.30	-1.64	0.11
11-20	0.26	0.35	0.62	0.09	0.28		3.82	2.54	3.04	-1.28	0.50
21-30	0.16	0.27	0.47	0.11	0.20		5.11	2.99	2.62	-2.12	-0.37
31+	0.42	0.49	0.88	0.07	0.39		3.50	3.01	2.52	-0.50	-0.49
EDU=5											
0-10	0.35	0.47	0.61	0.12	0.13		3.50	4.26	4.24	0.76	-0.03
11-20	0.20	0.34	0.46	0.14	0.13		6.20	7.25	6.77	1.05	-0.48
21-30	0.22	0.25	0.35	0.03	0.10		6.13	7.16	7.46	1.03	0.30
31+	0.23	0.28	0.56	0.05	0.28		2.41	4.82	5.85	2.42	1.03
EDU=6											
0-10	0.39	0.42	0.55	0.03	0.13		2.12	2.77	3.87	0.65	1.10
11-20	0.22	0.36	0.54	0.14	0.18		2.33	5.42	5.98	3.09	0.56
21-30	0.14	0.29	0.38	0.15	0.09		3.46	2.71	5.22	-0.75	2.51
31+	0.38	0.46	0.47	0.08	0.01		2.64	1.91	2.32	-0.74	0.42

Table 4 Within-group variance and composition change, female

	Within-group variance			change			Labor force share			change	
	1995	**2002**	**2007**	**95-02**	**02-07**		**1995**	**2002**	**2007**	**95-02**	**02-07**
EDU=1											
0-10	0.54	0.82	0.46	0.29	-0.36		1.00	0.66	0.42	-0.34	-0.23
11-20	0.88	0.55	1.14	-0.33	0.59		1.81	0.54	0.41	-1.26	-0.13
21-30	0.86	0.27	1.63	-0.59	1.36		2.77	0.89	0.84	-1.87	-0.06
31+	2.58	0.55	1.67	-2.03	1.12		0.58	0.28	0.65	-0.30	0.37
EDU=2											
0-10	0.70	0.67	0.86	-0.03	0.19		5.68	3.27	2.59	-2.41	-0.68
11-20	0.49	0.62	0.67	0.13	0.05		9.92	6.74	4.25	-3.19	-2.49
21-30	0.54	0.45	0.91	-0.09	0.46		14.69	8.24	6.68	-6.45	-1.56
31+	1.68	0.41	2.77	-1.27	2.36		2.10	2.66	4.83	0.56	2.17
EDU=3											
0-10	0.58	0.62	0.68	0.04	0.06		6.42	4.36	3.03	-2.06	-1.33
11-20	0.34	0.54	0.68	0.19	0.14		13.51	9.44	6.84	-4.07	-2.61
21-30	0.35	0.44	0.81	0.09	0.37		5.56	14.67	12.97	9.12	-1.71
31+	3.38	0.41	1.48	-2.97	1.07		0.68	1.72	4.70	1.04	2.98
EDU=4											
0-10	0.71	0.62	0.67	-0.09	0.06		4.87	4.07	3.02	-0.80	-1.06
11-20	0.30	0.48	0.43	0.18	-0.05		5.07	4.10	3.99	-0.97	-0.11
21-30	0.34	0.38	0.70	0.03	0.32		5.63	4.95	3.79	-0.68	-1.15
31+	1.32	0.70	2.17	-0.63	1.47		1.98	2.03	2.38	0.04	0.35
EDU=5											
0-10	0.36	0.55	0.73	0.18	0.19		3.23	6.52	6.65	3.30	0.13
11-20	0.23	0.30	0.41	0.06	0.12		4.52	8.60	9.36	4.07	0.76
21-30	0.17	0.29	0.59	0.12	0.30		4.07	6.05	7.56	1.99	1.51
31+	0.49	0.33	1.01	-0.16	0.68		0.75	2.21	3.02	1.46	0.80
EDU=6											
0-10	0.43	0.56	0.57	0.13	0.01		1.74	2.54	4.47	0.81	1.93
11-20	0.15	0.66	0.40	0.51	-0.26		1.33	3.46	4.50	2.13	1.04
21-30	0.17	0.23	0.33	0.06	0.10		1.51	1.53	2.22	0.02	0.69
31+	0.59	0.36	0.78	-0.22	0.42		0.60	0.45	0.84	-0.15	0.39

Table 5 Probit Models

	In year 2002=1/otherwise=0			
	Male		Female	
	1995 vs 2002	2002 vs 2007	1995 vs 2002	2002 vs 2007
Middle school	-0.654**	0.022	-0.080	-0.161
	(0.297)	(0.284)	(0.154)	(0.183)
High school	-0.344	0.202	0.019	-0.069
	(0.293)	(0.277)	(0.151)	(0.180)
Technical school	-0.504*	-0.059	0.146	-0.128
	(0.296)	(0.280)	(0.153)	(0.180)
3 year college	-0.036	-0.027	0.697***	-0.322*
	(0.294)	(0.275)	(0.153)	(0.175)
4 year college and above	0.008	-0.233	0.492***	-0.642***
	(0.298)	(0.277)	(0.166)	(0.181)
Experience_11-20	-0.551	0.105	-0.450**	-0.087
	(0.345)	(0.349)	(0.192)	(0.247)
Experience_21-30	-0.689**	0.135	-0.417**	-0.250
	(0.307)	(0.302)	(0.175)	(0.214)
Experience 30+	-0.391	0.035	-0.129	-0.722***
	(0.303)	(0.288)	(0.249)	(0.259)
MiddleXexper11-20	0.872**	0.120	0.548***	0.232
	(0.356)	(0.364)	(0.207)	(0.263)
MiddleXexper21-30	0.948***	0.016	0.400**	0.261
	(0.318)	(0.317)	(0.189)	(0.231)
Middle Xexper31+	1.248***	-0.141	0.656**	0.408
	(0.315)	(0.303)	(0.270)	(0.277)
HighXexper11-20	0.541	-0.156	0.465**	0.047
	(0.351)	(0.357)	(0.204)	(0.259)
HighXexper21-30	1.363***	-0.254	1.260***	0.108
	(0.315)	(0.310)	(0.188)	(0.226)
High Xexper31+	0.831***	-0.656**	0.970***	-0.035
	(0.315)	(0.298)	(0.283)	(0.277)
TechnicalXexper11-20	0.641*	-0.215	0.429**	-0.097
	(0.358)	(0.363)	(0.210)	(0.262)
TechnicalXexper21-30	0.699**	-0.063	0.447**	0.229
	(0.320)	(0.318)	(0.193)	(0.231)
TechnicalXexper31+	0.650**	0.116	0.270	0.511*
	(0.317)	(0.304)	(0.271)	(0.280)
3yearcolXexper11-20	0.525	-0.106	0.410**	0.022
	(0.352)	(0.355)	(0.207)	(0.255)
3yearcolXexper21-30	0.663**	-0.204	0.222	0.108
	(0.316)	(0.309)	(0.193)	(0.224)
3yearcolXexper31+	0.704**	-0.151	0.354	0.583**
	(0.315)	(0.297)	(0.279)	(0.274)
4yearcolXexper11-20	0.915**	0.009	0.810***	0.245
	(0.358)	(0.357)	(0.228)	(0.263)
4yearcolXexper21-30	0.369	-0.359	0.188	0.329
	(0.323)	(0.313)	(0.221)	(0.242)
4yearcol Xexper31+	0.025	0.041	-0.269	0.686**
	(0.322)	(0.304)	(0.318)	(0.310)
Central	-0.020	-0.018	-0.033	-0.038
	(0.030)	(0.030)	(0.033)	(0.033)
Eastern	-0.030	-0.312***	-0.002	-0.310***
	(0.030)	(0.028)	(0.032)	(0.031)
Constant	0.077	-0.072	-0.428***	0.153
	(0.288)	(0.270)	(0.143)	(0.170)
Pseudo R2	0.039	0.025	0.057	0.027
N	11607	13409	9912	11025

Table 6 Decomposition results

	Male			Female		
	Price of unobserved skill of			Price of unobserved skill of		
	1995	2002	2007	1995	2002	2007
A: Residual I						
quant90-10						
1995	1.14	1.41	1.57	1.32	1.47	1.74
2002	1.11	1.37	1.57	1.28	1.44	1.73
2007	1.12	1.38	1.58	1.30	1.47	1.80
quant50-10						
1995	0.60	0.80	0.85	0.75	0.84	0.96
2002	0.57	0.77	0.86	0.71	0.81	0.95
2007	0.57	0.76	0.86	0.72	0.83	1.00
quant90-50						
1995	0.54	0.61	0.72	0.57	0.63	0.78
2002	0.53	0.60	0.72	0.57	0.63	0.78
2007	0.55	0.61	0.73	0.58	0.64	0.81
Variance						
1995	0.33	0.46	0.58	0.51	0.48	0.76
2002	0.32	0.42	0.60	0.48	0.46	0.74
2007	0.33	0.40	0.61	0.57	0.45	0.83
B: Residual II						
quant90-10						
1995	1.11	1.28	1.38	1.23	1.40	1.52
2002	1.16	1.29	1.40	1.17	1.38	1.50
2007	1.14	1.36	1.45	1.16	1.43	1.50
quant50-10						
1995	0.57	0.69	0.73	0.69	0.76	0.70
2002	0.61	0.70	0.73	0.65	0.76	0.77
2007	0.79	0.73	0.75	0.73	0.79	0.78
quant90-50						
1995	0.53	0.58	0.65	0.54	0.64	0.82
2002	0.55	0.59	0.67	0.52	0.62	0.73
2007	0.34	0.63	0.70	0.43	0.64	0.71
Variance						
1995	0.31	0.34	0.34	0.42	0.37	0.41
2002	0.63	0.38	0.40	0.33	0.40	0.43
2007	0.57	0.40	0.46	0.26	0.42	0.47

Table 7 Explain the rising residual inequality

	Dependent=residual inequality in terms of								
	Variance			quant90-50			Quant50-10		
	95-07	95-02	02-07	95-07	95-02	02-07	95-07	95-02	02-07
A: Male									
soe	-0.509**	-0.720***	0.226	-0.403***	-0.301***	-0.093	-1.015***	-1.380***	-0.788
	(0.216)	(0.236)	(0.652)	(0.103)	(0.089)	(0.267)	(0.298)	(0.461)	(0.918)
College share	0.098	-0.126	0.365	0.024	0.318	-0.398	0.456	1.193	0.793
	(0.502)	(0.530)	(0.890)	(0.239)	(0.201)	(0.365)	(0.691)	(1.035)	(1.253)
Exp/gdp	1.296**	-0.134	2.169**	0.424*	0.340	0.886**	-0.089	0.352	0.003
	(0.489)	(0.780)	(0.953)	(0.233)	(0.296)	(0.390)	(0.673)	(1.524)	(1.342)
Constant	0.465*	0.937***	-0.154	0.719***	0.563***	0.628***	1.264***	1.263**	1.018
	(0.272)	(0.276)	(0.531)	(0.129)	(0.105)	(0.218)	(0.374)	(0.539)	(0.748)
R2_W	0.219	0.294	0.180	0.304	0.408	0.196	0.237	0.358	0.051
R2_B	0.001	0.027	0.000	0.083	0.010	0.066	0.014	0.001	0.002
R2_O	0.015	0.116	0.005	0.141	0.066	0.060	0.158	0.107	0.008
N	99	66	66	99	66	66	99	66	66
B: Female									
soe	-0.321	-0.143	-0.407	-0.503***	-0.273**	-0.423	-0.604*	-0.418	-0.326
	(0.281)	(0.350)	(0.886)	(0.122)	(0.103)	(0.413)	(0.325)	(0.314)	(1.067)
College share	1.047	1.254	1.119	0.325	0.449*	0.449	0.208	0.680	0.136
	(0.649)	(0.872)	(1.307)	(0.281)	(0.257)	(0.609)	(0.752)	(0.783)	(1.573)
Exp/gdp	2.008***	1.255	2.365*	0.295	0.231	0.349	1.740**	-0.226	2.843*
	(0.616)	(1.216)	(1.212)	(0.267)	(0.358)	(0.565)	(0.714)	(1.093)	(1.460)
Constant	0.086	0.066	0.023	0.720***	0.555***	0.650**	0.749**	0.897***	0.464
	(0.272)	(0.349)	(0.588)	(0.118)	(0.103)	(0.274)	(0.315)	(0.314)	(0.707)
R2_W	0.293	0.125	0.273	0.435	0.431	0.161	0.198	0.182	0.185
R2_B	0.003	0.000	0.017	0.182	0.303	0.109	0.000	0.009	0.000
R2_O	0.035	0.003	0.037	0.330	0.358	0.116	0.016	0.067	0.006
N	99	66	66	99	66	66	99	66	66
C: Male									
soe	-0.343*	-0.581***	0.508	-0.331***	-0.450***	0.153	-0.575***	-0.807***	0.031
Exp/gdp	1.306***	-0.121	2.157***	0.420*	0.033	0.714	0.251	-0.096	0.784
D: Female									
soe	-0.077	-0.028	0.135	-0.325**	-0.308**	0.098	-0.335**	-0.271	0.209
Exp/gdp	1.705***	0.737	2.084***	0.463	0.111	0.800	0.823**	-0.454	1.574**

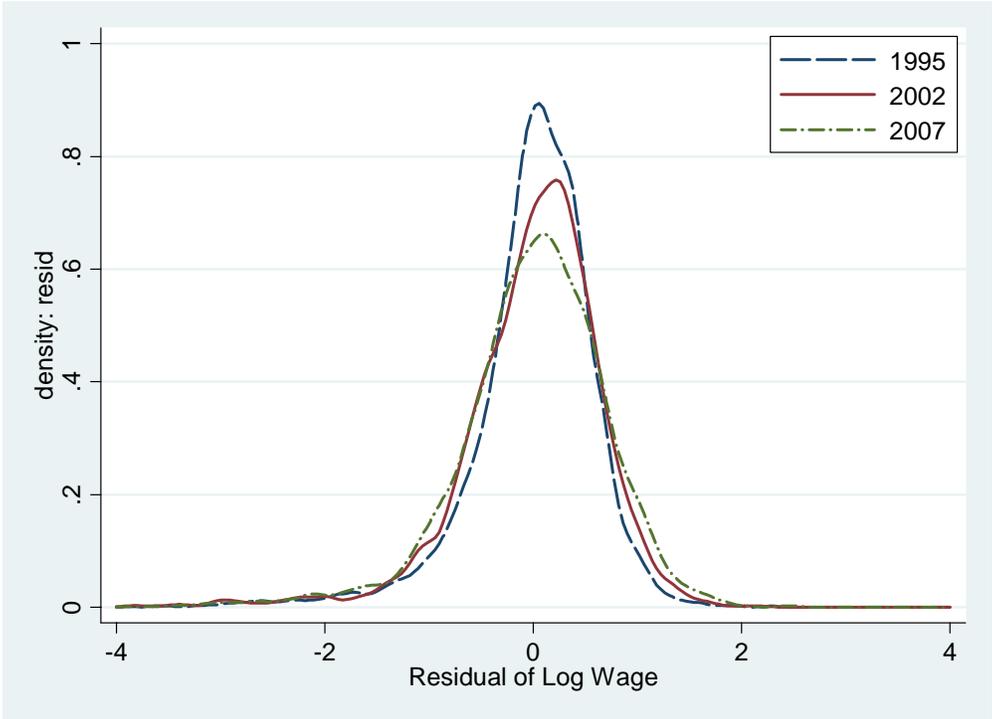Figure 1a Residual Ln Wage Distributions, male

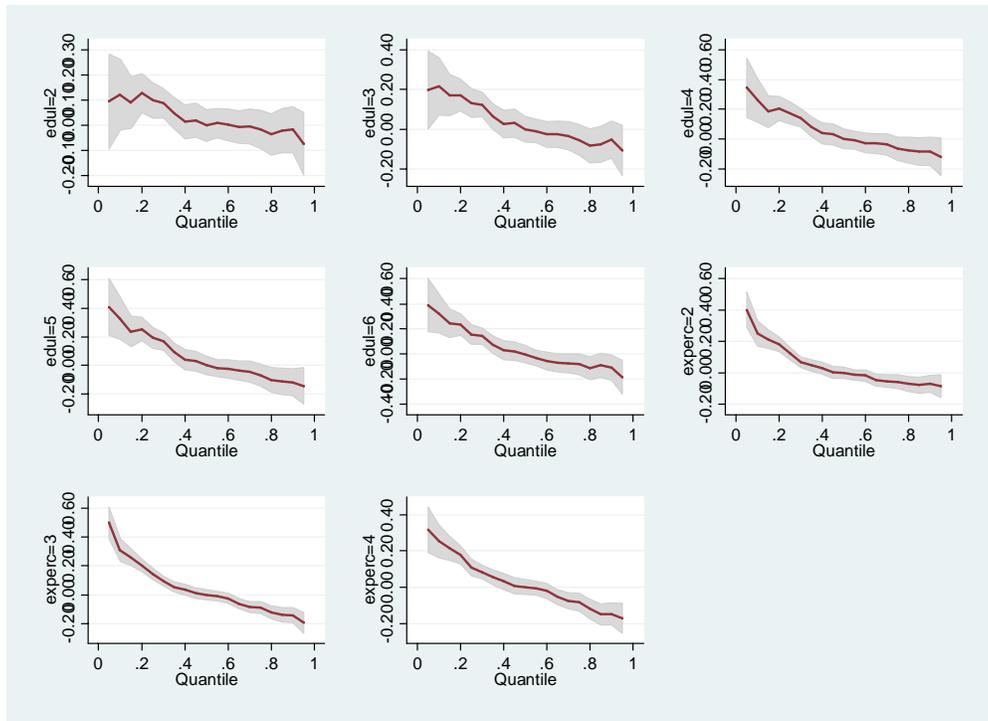Figure 1b Residual Ln Wage Distributions, female

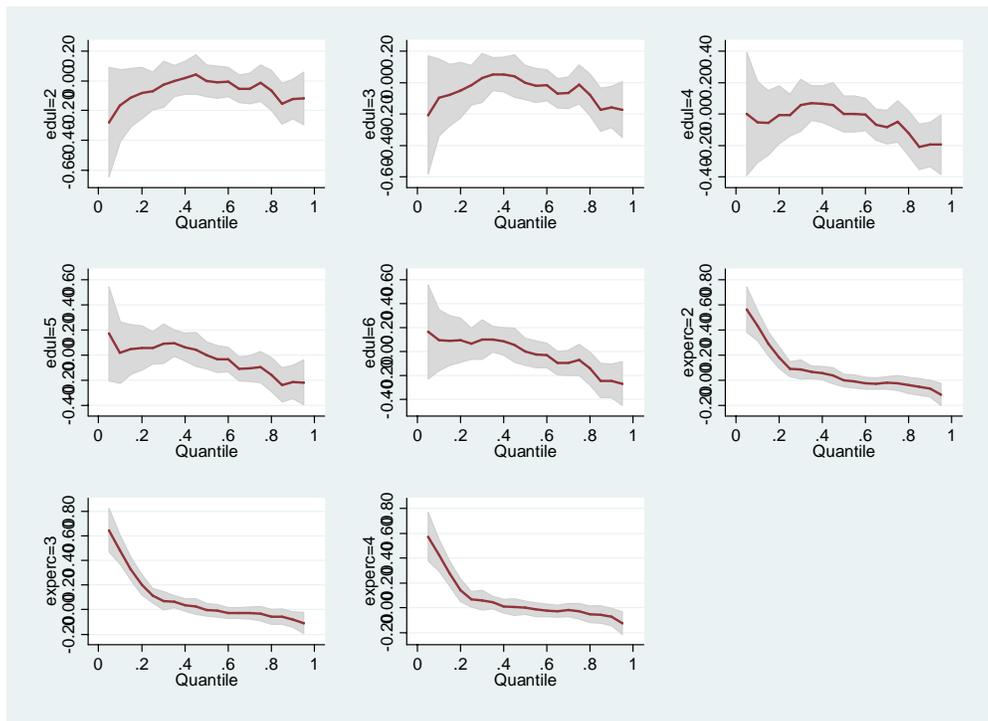Figure 2a Coefficients in quantile regressions, 1995 male

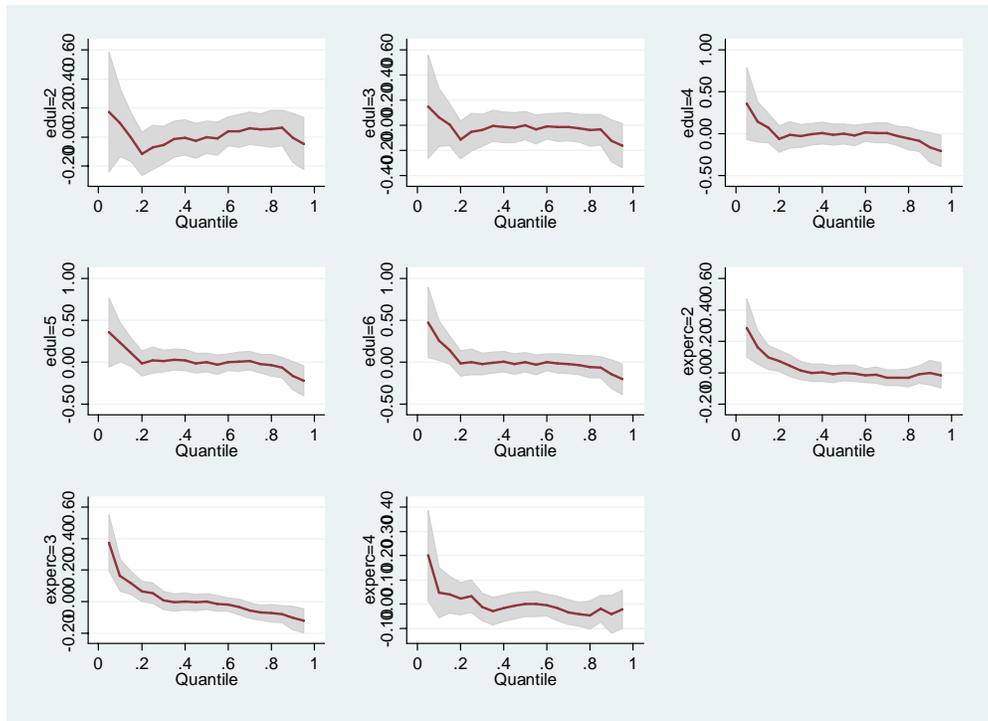Figure 2b Coefficients in quantile regressions, 2002 male

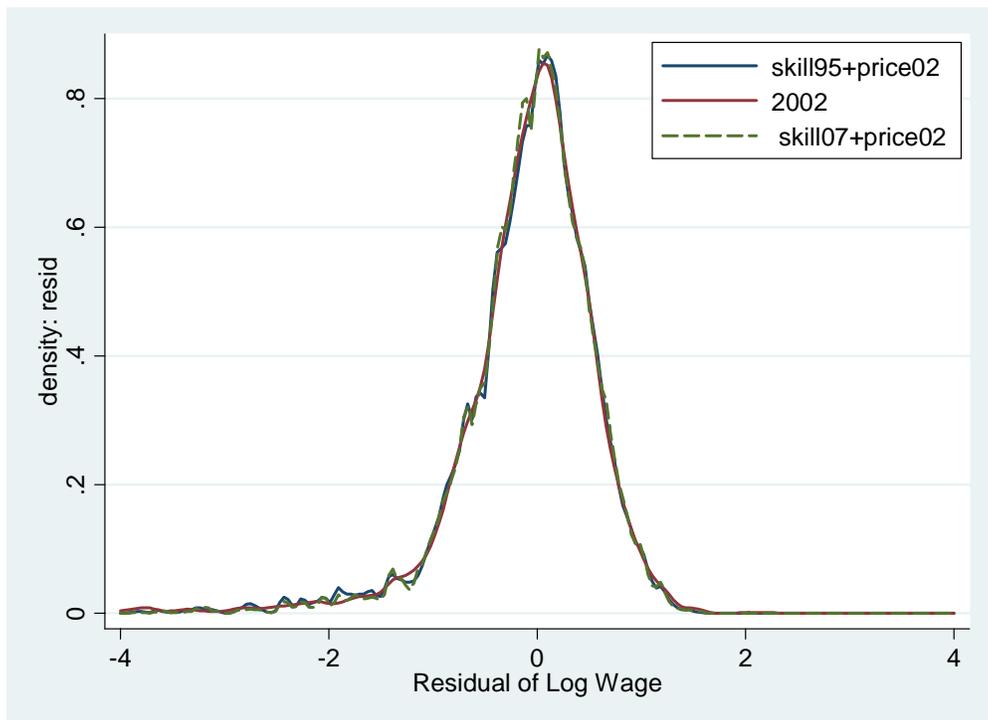Figure 2c Coefficients in quantile regressions, 2007 male

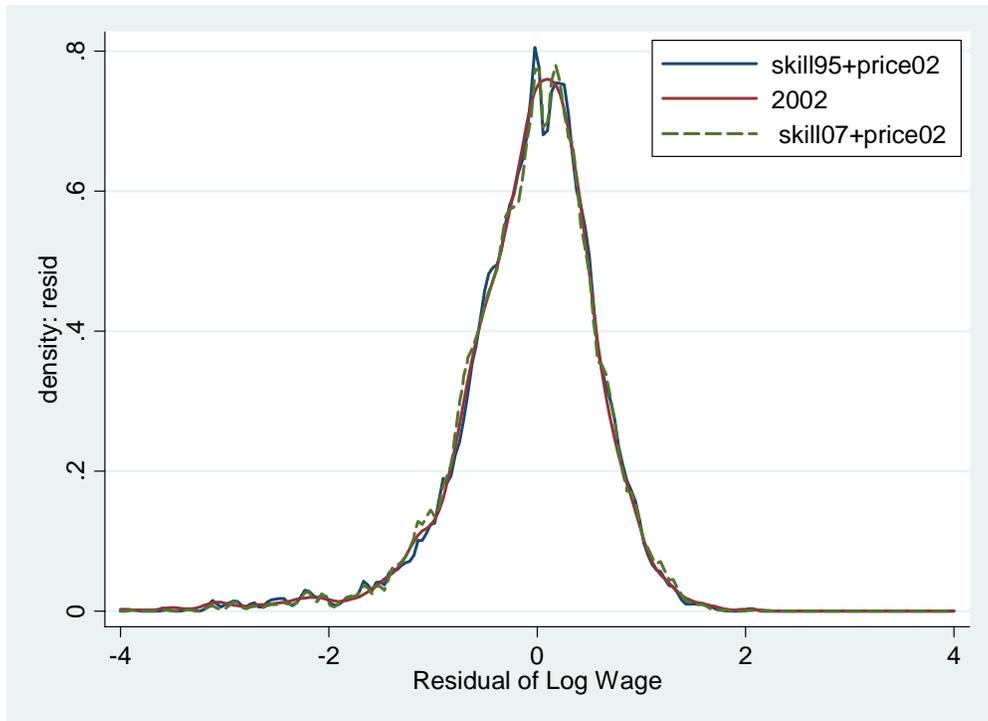Figure 3a Factual and Counterfactual Residual Distributions (MM), male

Figure 3b Factual and Counterfactual Residual Distributions (MM), female
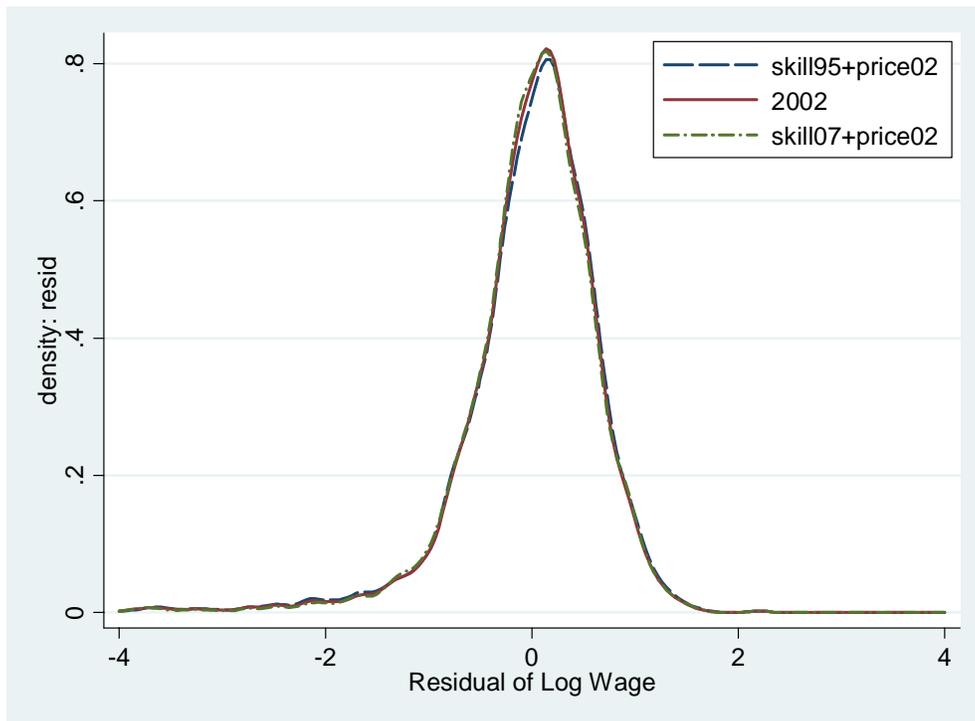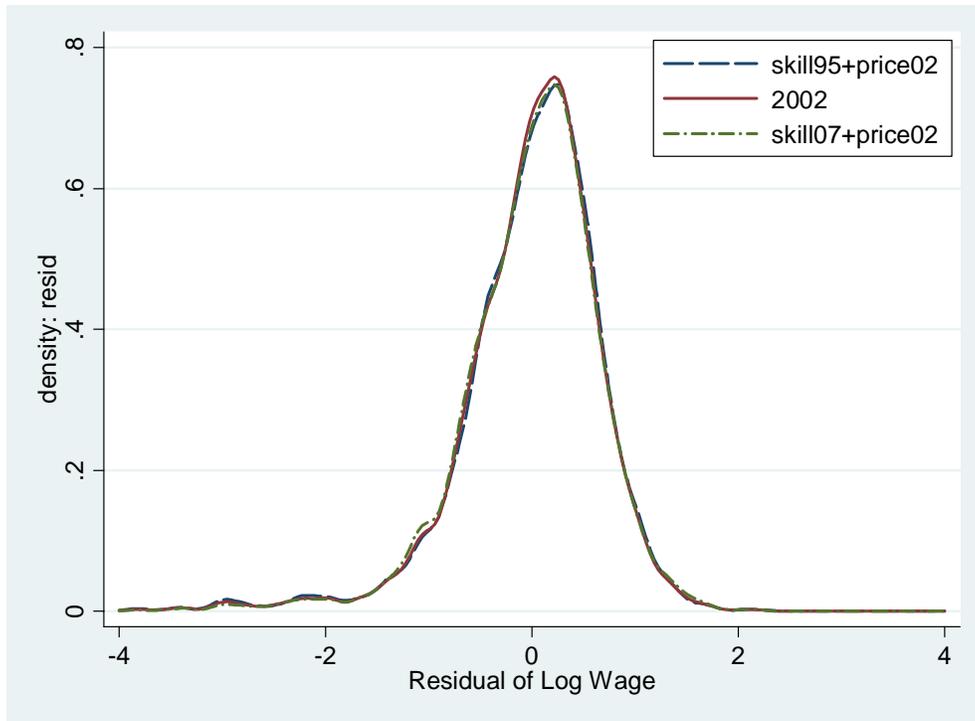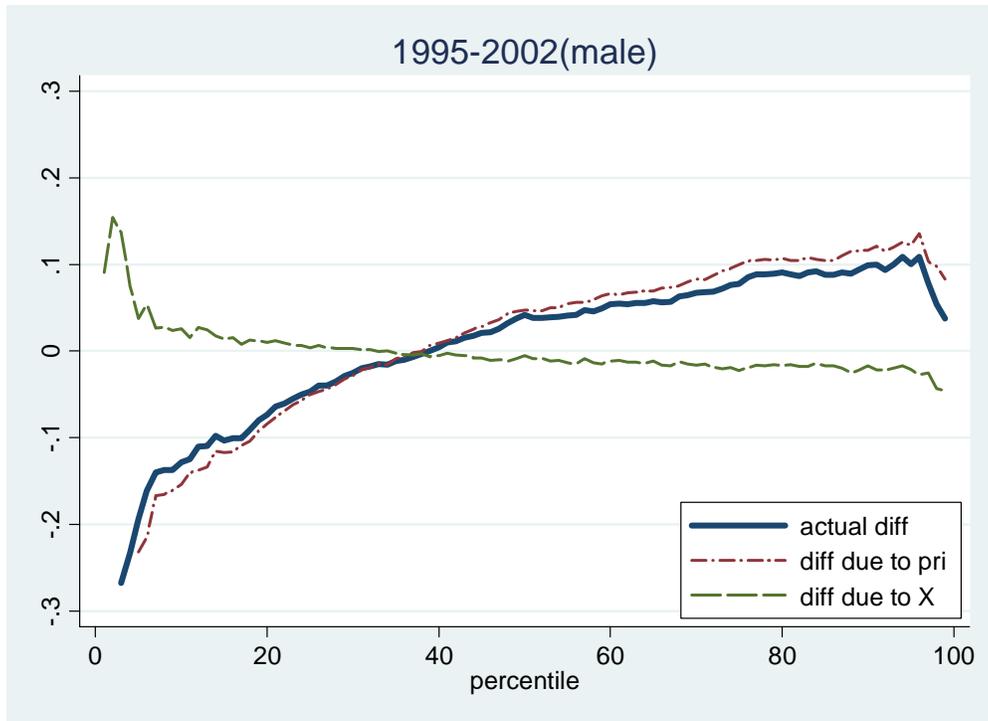

Figure 4a Factual and Counterfactual Residual Distributions (DFL), male

Figure 4b Factual and Counterfactual Residual Distributions (DFL), female

Figure 5a Decomposing Change in Residual Wage Distributions
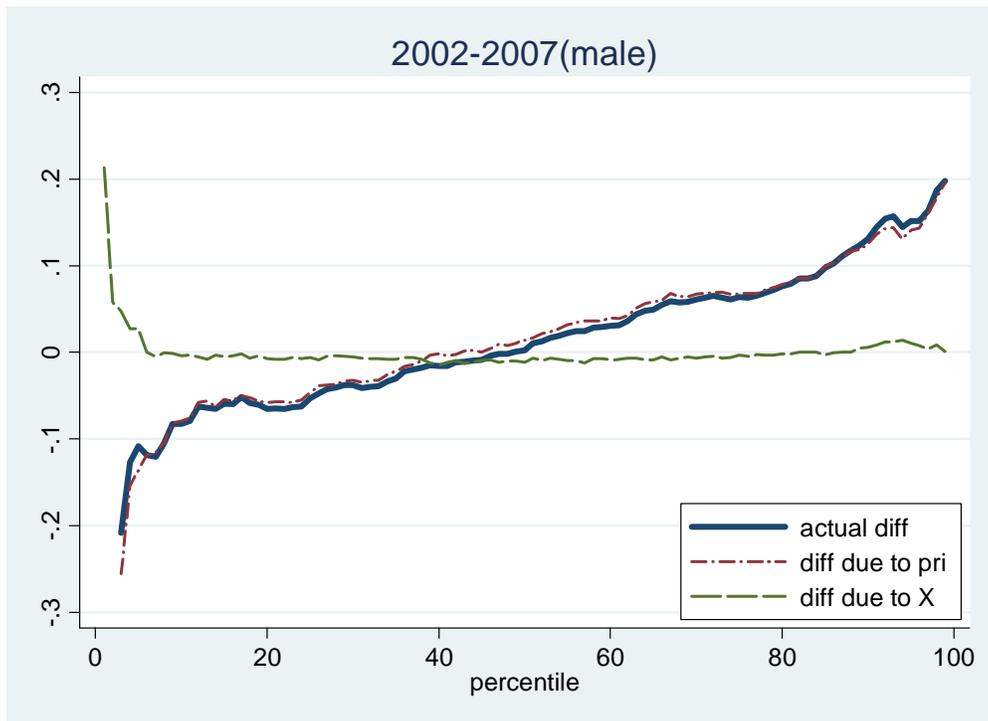
Figure 5b Decomposing Change in Residual Wage Distributions
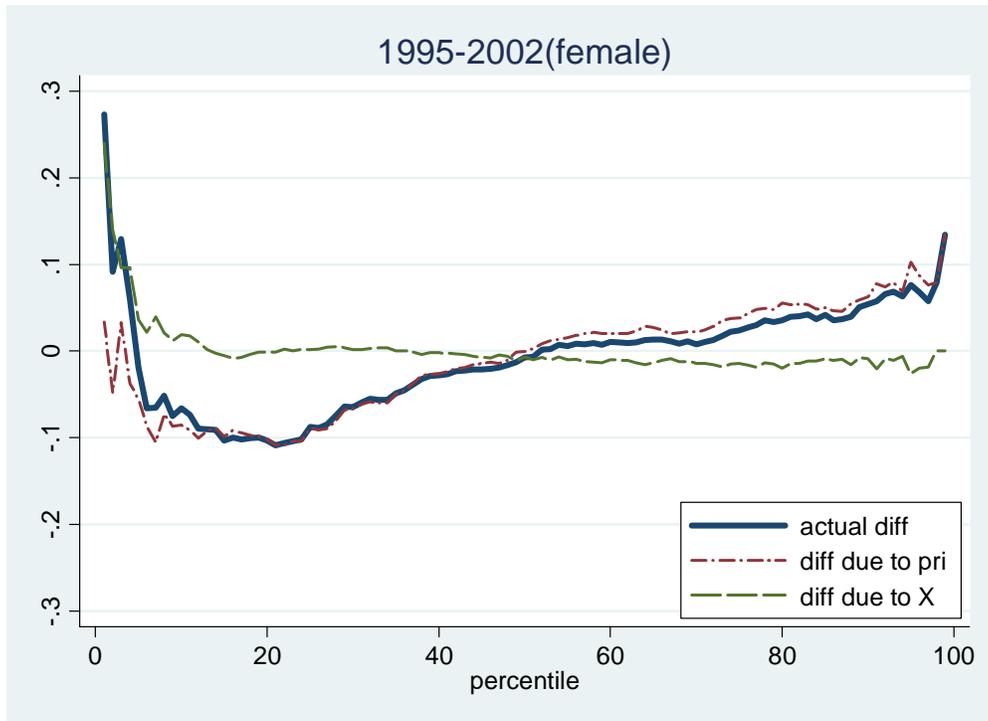
Figure 5c Decomposing Change in Residual Wage Distributions
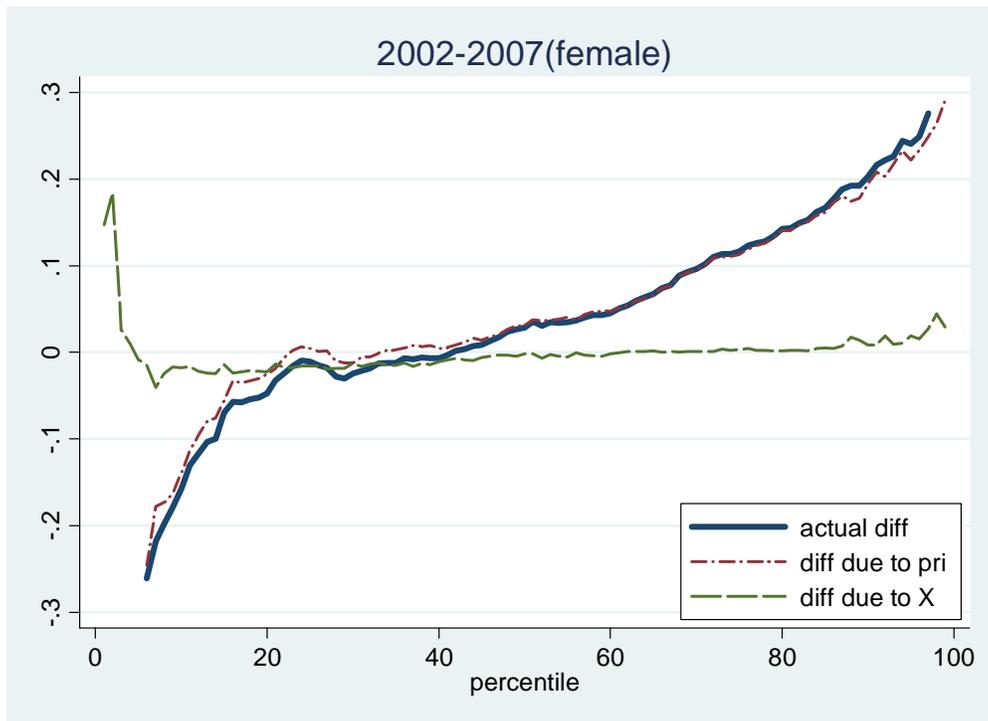
Figure 5d Decomposing Change in Residual Wage Distributions
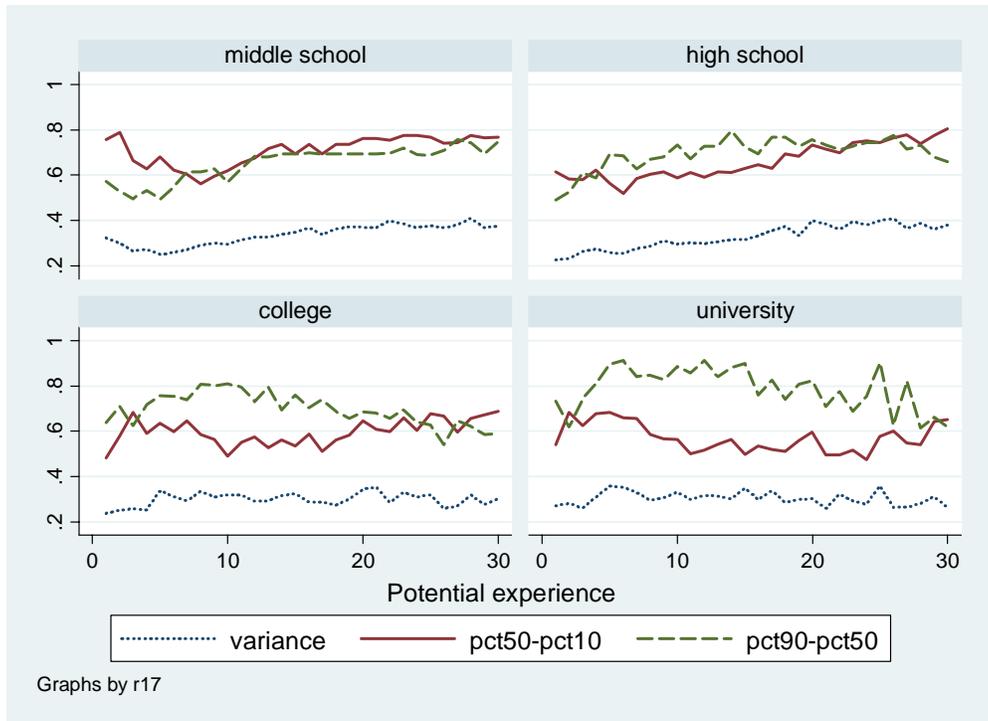
Figure 6a Within-Group Inequality for Male, 2005
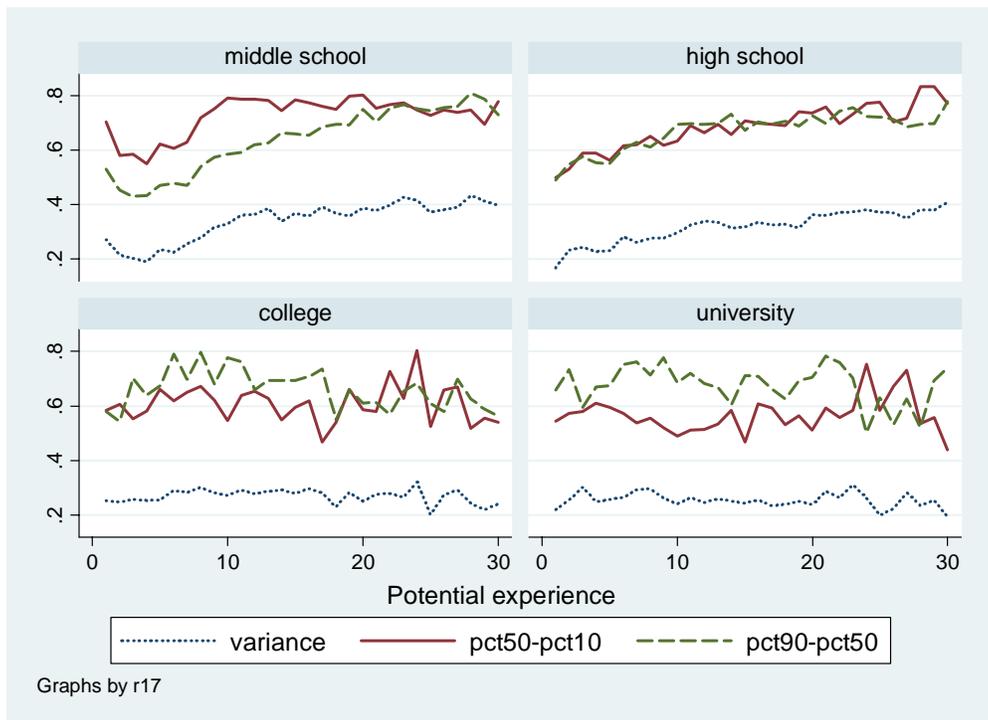
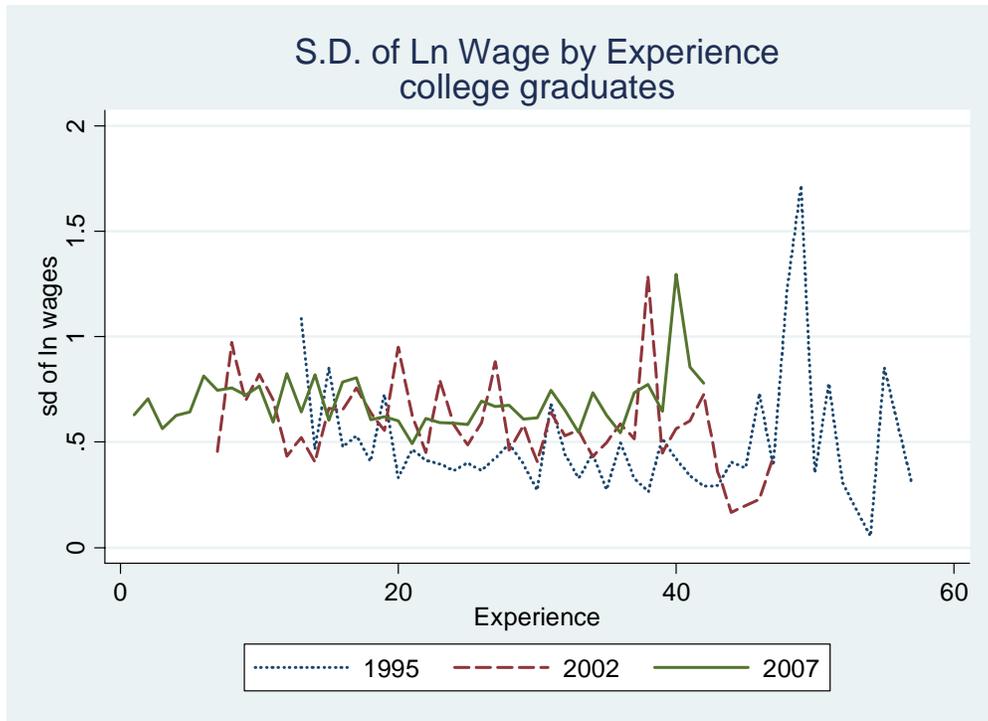Figure 6b Within-Group Inequality for Female, 2005

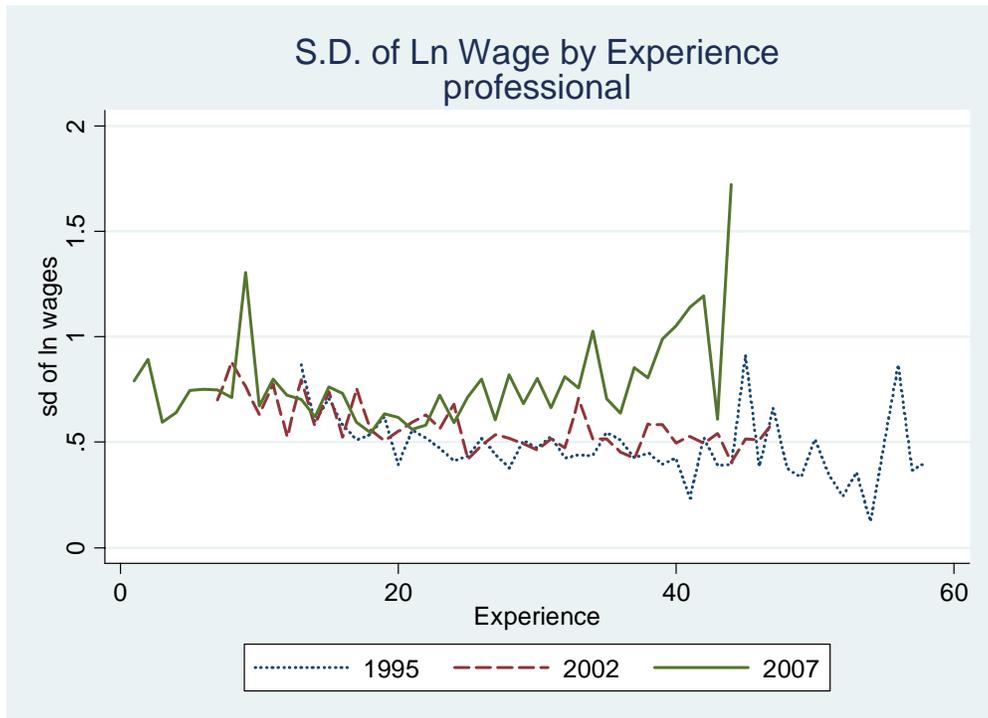Figure 7a Within-group Inequalities in Three CHIP Surveys (College Graduates)

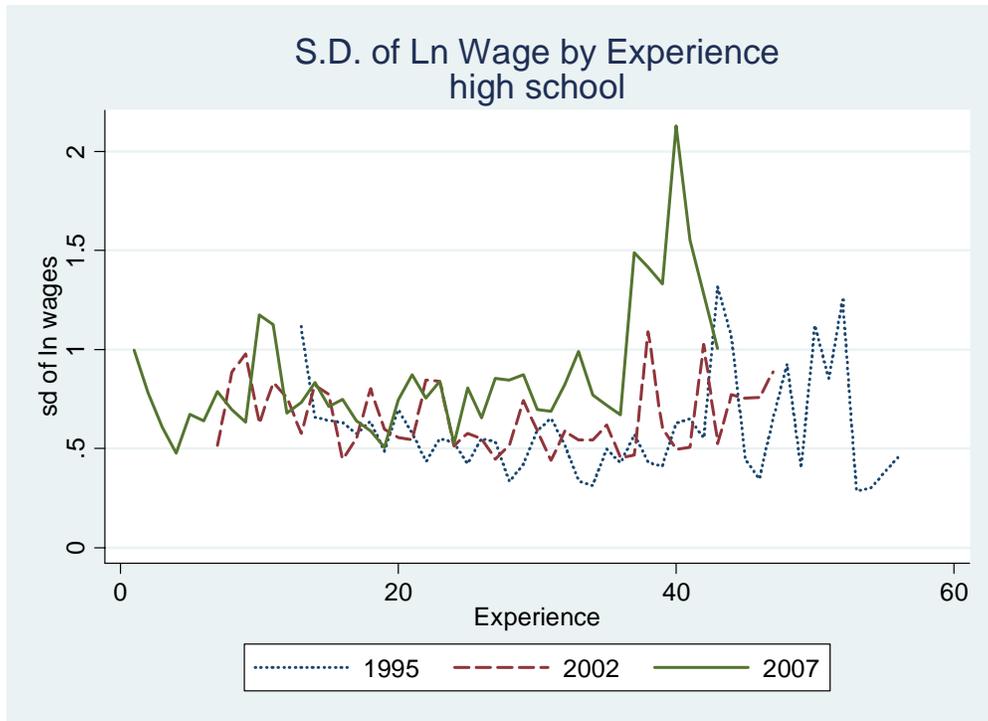Figure 7b Within-group Inequalities in Three CHIP Surveys (Professional)

Figure 7c Within-group Inequalities in Three CHIP Surveys (High School Graduates)