ANALYSIS OF CHANGES IN EARNING DISTRIBUTIONS OF URBAN CHINESE ECONOMY USING QUANTILE REGRESSION

WANG ZIJUN

(MASTER OF SOCIAL SCIENCES, NUS)

A THESIS SUBMITTED

FOR THE DEGREE OF MASTER OF SOCIAL SCIENCES

DEPARTMENT OF ECONOMICS

NATIONAL UNIVERSITY OF SINGAPORE

2010

Acknowledgement

I am heartily grateful to Professor Chen Songnian, my supervisor, whose patient instructions and continuous encouragement throughout the whole academic year helped me to understand the topic and enabled me to develop this thesis.

In addition, I wish to express my sincere thanks to A/Professor Liu Haoming, who has given me a lot of valuable suggestions for improving this thesis.

Table of Contents

Summary

Increased wage inequality has been observed in many countries. The chief explanation is that the increasing demand for highly skilled labor, which seems to be caused by the spread usage of computers, raises wages for high-skilled workers. This paper studies the change in earning distributions of the urban Chinese economy, with quantile regression and counterfactual decomposition analysis. Specifically, we examine how the wage distribution has changed in urban areas of China between 1995 and 2002; furthermore, we decompose the changes to be the consequence of workers' characteristics changes and the consequence of changes in rates of return to these characteristics. We find that both real wage and real wage inequality in urban areas of China have increased significantly during the period. Our model, which displays high accuracy in estimating the real wage distributions, shows that both gender gap and rates of return to education has increased between the two years. With counterfactual decomposition analysis, we find that changes in gender gap and return to education have contributed most toward the increased wage inequality.

List of Tables

List of Figures

1. Introduction

In the last decade, increased wage inequality has been observed in many countries. A chief explanation for the larger wage inequality is that because of the widely spread usage of computers and high technical machineries, the demand for high-skilled workers has been increasing; higher demand for high-skilled workers raises the relative wage of those skilled. This in return stimulated more people to pursue for higher education qualifications. In consequence, there was more supply of educated workers in the last decade than ever before, which further crowded out those unskilled, making the wage inequality yet more severe.

The economic reform of China in the 1970s has brought great changes to Chinese economy. The growth of the economy was accelerated after Deng Xiaoping's southern tour in 1992. For example, between 1995 and 2002, nominal GDP, as indicated by *China Statistical Yearbook*, approximately doubled. A notable change after the reform is the wage system. Before the economic reform, wages were set by some non-market mechanism, and studies have shown that the rate of return to education was quite low (*Gustafsson and Li, 2001*). But after the economic reform, Chinese economy became more market-oriented, so that people could be paid according to their qualifications, like working experiences and education levels.

The conventional way to study wage distributions is by employing the Mincerian wage equation, in which the role of education is always of the most interest. The equation is often estimated with OLS approach. However, the information that ordinary linear regression model could provide is too limited. Recently, many researches were done with quantile regression instead. For example, researches using quantile regression on Portugal data (*J.A.F. Machado and J. Mata, 2005*) have shown that low paid jobs and high paid jobs are paying different returns to the same level of education; and they also found that there is not only an increased wage inequality in the country, but also a more serious wage inequality within the skilled group of the country, which suggests that education itself is also a reason for increased wage inequality in Portugal.

Many papers are focusing on gender gap problem in China, for example, John A.Bishop, Feijun Luo and Fang Wang's paper in 2004 used quantile regression to identify the change of gender gaps in China between 1988 and 1995. They have found that low paid jobs display greater discrimination than high paid jobs, and that gender gap in 1995 is smaller than gender gap in 1988.

In this paper, we want to examine how the wage distribution has changed in urban Chinese economy after 1992, specifically, between 1995 and 2002. Furthermore, if there is indeed a change, we want to find out the reasons that have caused that change. In particular, we want to see whether the change in wage distributions is contributed by the overall changes of the work force characteristics, or by changes in rates of return to some characteristics.

The paper proceeds as follows. In Section2, we present the econometric models which we will use in the empirical part. Detailed information about our data is provided in Section3. Results with complete analysis are discussed in Section4 and Section5 concludes. Figures (all of them are produced with software "R") and tables are presented in the Appendix.

2. Modeling

2.1 Ordinary Least Squares

Linear regression is mostly used in modeling and analyzing the relationship between a response variable (denoted as *Y*) and p explanatory variables (denoted as a $p \times 1$ vector *X*):

$$E(Y \mid X) = X\beta$$

For convenience, we usually write the model as

$$Y = X\beta + u \ ,$$

where u is the error term, assumed to have a mean of zero; and the model is assumed to be linearly dependent on the unknown parameter β which are to be estimated. The most popular way to solve for the unknown parameter is through the ordinary least squares (OLS) approach, i.e.

$$\min_{\beta} \sum_{i=1}^{n} (y_i - x_i' \beta)^2 \ .$$

A well-known attractive feature of OLS is that it provides the smallest mean-squared error linear estimation to the conditional mean function, regardless of whether the model is correctly specified or not.

Suppose we are interested in finding out the distribution of wages in some country, knowing only the mean of the national wages is far from enough. But if we know more information of the national wages, say, if we know the *10th* quantile, the *25th* quantile (which is the *1st* quarter), the *50th* quantile (which is the median), the *75th* quantile (which is the *3rd* quarter) and the *90th* quantile of the national wages, we should expect to see a bigger picture of the wage distribution than just from the mean. Likewise, if we are interested in finding out how some *X* is explaining *Y*, knowing only the information of the conditional means of *Y* given *X* is far from enough. Given some value of *X* (e.g. height), there could be a range of possible values of *Y* (e.g. weight), and therefore, given *X*, there is a conditional distribution of *Y*. If we have information on different conditional quantiles of *Y* given *X*, we could see a more complete picture of how *X* is affecting *Y*.

Introduced by Koenker and Bassett (1978), quantile regression fits a linear model for the conditional quantiles of the response variable, from which we are able to capture a bigger picture of how *X* is explaining *Y*.

2.2 Quantile Regression

Suppose that the conditional distribution function of *Y* given *X* is denoted as $F_Y(y \mid X)$, and that the conditional density is $f_Y(y \mid X)$. Let $\tau \in [0,1]$ to be such that

$$F_Y(y \mid X) = \tau \, .$$

Then, we could define the τth conditional quantile of *Y* given *X* to be

$$Q_\tau(Y \mid X) := \inf\{y : F_Y(y \mid X) \geq \tau\} \, .$$

The linear quantile regression models the τth conditional quantile of *Y* to be

$$Q_\tau(Y \mid X) = X\beta(\tau) \, . \qquad (*)$$

In the linear quantile regression model (*) above, *X* is the vector of covariates, as it used to be in the ordinary linear regression model, and $\beta(\tau)$ is the vector of coefficients that are of interests at the τth conditional quantile of *Y*; one has to note that in quantile regression model, at different quantiles of *Y*, we will have different estimates of the parameter vector.

The difference between linear quantile regression and ordinary linear regression is that we are fitting the conditional quantiles of *Y* given *X*, rather than just fitting the conditional means of *Y*. Just as quantiles capture more details than simply the mean, quantile regression could capture more details than ordinary linear regression.

The most familiar quantile to us is the median. For example, when we say a person has the median wage out of a population, we mean half of the population has lower wages than him, and the other half has greater wages than him. A well-known feature about the sample median of *Y* is that it solves

$$\min_{m} \sum_{i=1}^{n} | y_i - m | .$$

Correspondingly, the conditional median of *Y* given *X* solves

$$\min_{q(X)} \sum_{i=1}^{n} | y_i - q(X) | .$$

In the same way, the conditional quantile of *Y* solves

$$\min_{q(X)} \sum_{i=1}^{n} \rho_\tau (y_i - q(X)) ,$$

where $\rho_\tau(u) = (\tau - 1(u \leq 0))u$, and it is sometimes referred to as the loss function. Given the linear form of $Q_\tau(Y \mid X)$, those parameters of interest could be estimated by solving

$$\min_{\beta} \sum_{i=1}^{n} \rho_\tau (y_i - x_i^{\prime} \beta) .$$

2.3 Counterfactual Decomposition

As we have mentioned in the introduction section, our interest lies not only in how the wage distribution has changed over the years, but we are also interested in finding out what exactly has caused the change. For this purpose, we need to decompose the change in wage distributions into two parts: i) the consequence of changes in rates of return to different human capital characteristics over the years, and ii) the consequence of changes in the overall worker force characteristics over the years. To be more specific, we want to find out what would 2002 wage distribution like if the work force characteristics were as in 1995. Mathematically, denote $f(y^*(2002);x(2002))$ to be the wage density in 2002 if all the covariates are as in 2002 (with 2002 rates of return), i.e. density of

$$Q^*_\tau(Y \mid X)_{2002} \equiv X_{2002}\,\beta(\tau)_{2002} .$$

Furthermore, denote $f(y^*(2002);x(1995))$ to be the wage density in 2002 if all the covariates are as in 1995 (still with 2002 rates of return), i.e. density of

$$Q^*_\tau(Y \mid X)_{2002\ \textit{counterfac tual}} \equiv X_{1995}\,\beta(\tau)_{2002} .$$

Define $f(y^*(1995);x(1995))$ and $f(y^*(1995);x(2002))$ in the same way (both of which are with rates of return from 1995). Then we would have

$$\begin{array}{l}
\underbrace{f(y^*(2002);x(2002)) - f(y^*(1995);x(1995))}_{\Downarrow} \\
\underbrace{f(y^*(2002);x(2002)) - f(y^*(2002);x(1995))}_{\text{impact of covariates}} \\
+ \\
\underbrace{f(y^*(2002);x(1995)) - f(y^*(1995);x(1995))}_{\text{impact of coefficients}}
\end{array}$$

Or reversely (for the purpose of double check in the empirical part),

$$\begin{array}{l}
\underbrace{f(y^*(2002);x(2002)) - f(y^*(1995);x(1995))}_{\Downarrow} \\
\underbrace{f(y^*(2002);x(2002)) - f(y^*(1995);x(2002))}_{\text{impact of coefficients}} \\
+ \\
\underbrace{f(y^*(1995);x(2002)) - f(y^*(1995);x(1995))}_{\text{impact of covariates}}
\end{array}$$

Note that to see impact of coefficients, we look at the differences between wage densities that are estimated with the same set of work force characteristics but different years' rates of return (use coefficients from different years' models); on the other hand, to see the impact of covariates, we look at the differences between wage densities which are estimated with the same set of rates of return but different years' work force characteristics (use covariates from different years' datasets).

Clearly we need to estimate the so-called counterfactual densities $f(y*(2002);x(1995))$ and $f(y*(1995);x(2002))$. At the same time, we cannot use the unconditional marginal density of wages obtained directly from our datasets to be $f(y*(2002);x(2002))$ or $f(y*(1995);x(1995))$. This is because the marginal density of wages directly obtained from the data might not necessarily agree with our conditional model (*), which would serve as a basis for the model specification test later. Hence, we need to estimate all of the four densities listed above. The methodology we would follow comes from *J.A.F. Machado and J. Mata, 2005*:

Step1) randomly generate $\{\tau_i\}_{i=1}^{m}$ from the Uniform [0, 1] distribution.

Step2) estimate quantile regression coefficients $\{\beta(\tau_i)\}_{i=1}^{m}$ for the data set.

Step3) randomly generate a covariate sample $\{x_i\}_{i=1}^{m}$ with replacement from the dataset.

Step4) the estimated wage* - $\{y_i* \equiv x_i' \beta(\tau_i)\}_{i=1}^{m}$ - will have the marginal distribution $f(y*;x)$ that is consistent with the conditional model (*).

For example, $f(y*(2002);x(2002))$ will be estimated using 2002 dataset; and to estimate $f(y*(2002);x(1995))$, follow the steps above with 2002 dataset but generate the covariate sample in Step3 from the 1995 dataset instead. Similarly, we could estimate $f(y*(1995);x(1995))$ and $f(y*(1995);x(2002))$.

3. The Data and Our Model

3.1 Our Model

We employ the conventional Mincerian education equation, which is quite widely used when studying the impact of education on income. One thing worth mention is that there are still limitations on Mincerian education equation, e.g. there exist left out variables like capabilities of workers. However, those left out variables are most of the times difficult to measure, and hence would be taken as noises.

In this model, workers' characteristics are taken to be the covariates: gender (dummy variable, equal to 1 if female), years of education, years of potential experience (*by Mincer, 1974*) and potential experience square.

$$Q_\tau(y_i \mid X) = x_i^{'} \beta(\tau)$$

In the model above, y_i represents the natural logarithm of real wages for person *i* if he/she performs in the τth conditional quantile, with personal characteristics denoted as x_i. Potential experience is obtained by age minus years of education minus 7, where 7 refer to the age for entering primary school in China.

3.2 Data Source

The data used in this paper comes from the China Household Income Project, 1995 and 2002. The surveys were conducted in both rural and urban areas of China by the National Bureau of Statistics of China every seven years. We use the data from the urban surveys, and only include cities that were surveyed in both years, so that the two datasets would be more comparable. As a fact in China, women are required to retire after 55 years old while men are required to retire after 60; in order to avoid the noise on gender that would be raised by age, we restrict the observations to be those with ages above 18 (adults) and below 55. Moreover, we consider only the work force population without students or retirees, since they might be paid according to different policies or systems. And the samples we are interested in are those full-time employees out of the population. The reason we exclude those unemployed is that they don’t have wages data. Observations in both data sets are workers from industry sectors, including the government, manufacturing, health, education, services, trade, construction, communication or restaurant sectors, etc. There are 8180 observations in 1995 sample and 7347 observations in 2002 sample.

3.3 (log) Real Wage Descriptions

We will only look at log real wages (annually) in the context below, where 2002 is the base year. Here, real wages are calculated from the data, as the data provides Consumer Price Index for both years, where 2002 has CPI taken as 100.

Figure1 plots the unconditional density of log real wages in both years, with the blue dotted curve representing 1995 and the red curve representing 2002. Very clearly, the wage density curve shifts to the right from 1995 to 2002, indicating an overall increase in the real wage level. From the summary statistics in Table1 (providing the minimum, the *1st* quarter, the median, the *3rd* quarter, the maximum, the mean and the standard deviation of log real wages in both years), we see more clearly that all the quarters and mean of wages are higher in 2002, which is consistent with the density curves. Both density curves have just one mode with bell shaped appearances. If we look at the spread of the two curves, we can see 1995 curve is taller and a bit more centered around the mode, while 2002 curve is shorter and more spread out, which is evidence of higher wage inequality in 2002, compared with 1995. This is confirmed by the statistics in Table1. Clearly, the standard deviation of real wages in 2002 is larger than that of 1995, which means the real wage spread is larger in 2002 than in 1995, and hence shows a larger inequality.

3.4 Work Force Characteristics

Table2 provides some summary statistics of the work force characteristics, and Figure2 plots the density curves for the continuous variables (with blue dotted curve denotes 1995 and red curve denotes 2002).

Female workers made up 48.4% of our 1995 sample, while it decreased to 45.5% in 2002 sample. According to *China Statistical Yearbook,* female represents 48.97% of the total national population in 1995, which has felled to 48.47% in 2002. We have examined the percentage of female in the urban work force population (including both employed and unemployed aged between 18 and 55), and found that it was 49.5% in 1995 and 48.3% in 2002. At the same time, employment rate of female was 93.2% in 1995, whereas it was only 81.3% in 2002. Hence, not only percentage of female in work force but also female employment rate has decreased over the years. One explanation for this is that, as Chinese economy as a whole has advanced, the average family income has increased, so that fewer housewives need go out for jobs. However, we noticed that the decrease in female employment rate is much bigger than that in percentage of female in work force population; it’s possible that gender discrimination was more serious in 2002 labor market compared with 1995 labor market.

As the number of admissions to colleges and universities is increasing in China year after year, more people could have the chances to receive higher educations. Meanwhile, the market is becoming more and more competitive gradually; in order to be competitive candidates in the job market, people have to try and obtain higher education qualifications. Therefore, we could expect the general education level to be increased during the decade. The average number of years of education is 10.9 in 1995 and 11.6 in 2002. There are less number of workers who haven't finished the 9-year compulsory education and more people are with higher level of education in 2002. More specifically, in 2002, the percentage of workers with 9-12 (inclusive of 12) years of education is 2.4% higher; percentage with 12-16 (inclusive of 16) years of education is 5.8% higher and percentage with 16-24 (inclusive of 24) years of education is 0.8% higher. The first panel in Figure2 (in which the blue dotted curves represent 1995 and red curves represent 2002) shows that both 1995 and 2002 education curves have three modes. The two modes on the right are higher for 2002, indicating that in 2002, bigger fraction of workers is with more than 12 years of education. This also implies an increase in the supply of high-skilled workers.

Finally, both the summary statistics of potential experiences in Table2 and the second panel of Figure2 show that, the number of years of potential experience is larger for 2002 sample than that of 1995 sample. According to *China Statistical Yearbook*, the life expectancy has increased by approximately 3 years from 1990 to 2000. Probably because of much more advanced medical technology and health check plans, people could enjoy much healthier life and live longer. Hence, compared with 1995, in 2002 more workers could exit the market at older ages or until they are required to retire, resulting in higher potential experiences.

4. The Results and Discussion

4.1 Quantile Regression Estimates

With the model and data mentioned in the last section, we have plotted very comprehensive quantile regression estimates provided in Figure3, which is partitioned into 6 parts. Figure3.1 explains the estimations of the intercept, Figure3.2 provides gender coefficients, and Figure3.3 analyzes rates of return to education while Figure3.4-3.6 focus on effect of potential experience. As we have already mentioned before that for each conditional quantile τ of *Y*, we have one estimate for the coefficient vector $\beta(\tau)$. In the first row of each partition, coefficients are estimated from *1st*, *2nd* to *98th*, *99th* conditional quantiles and the estimates are plotted against the corresponding quantiles, with the left panel refers to year 1995 while the right panel refers to year 2002. The 95% confidence bands are plotted as blue bands. In addition, a horizontal red line denotes the OLS estimate of that coefficient in each panel, with the dashed red lines denoting the 95% confidence interval for OLS estimate. In the second row of each partition, we plot the change of coefficients (2002 coefficient value minus 1995 coefficient value) at *10th* to *90th* quantiles with 95% confidence intervals.

Table3.1 and Table3.2 list the detailed OLS estimates and the quantile regression estimates at some typical quantiles: *10th*, *25th*, *50th*, *75th* and *90th*, for year 1995 and 2002 respectively. The standard error of each estimate is also provided, with “*” denoting significant result at 5% significance level.

The intercept term refers to a male worker with zero years of education and zero years of potential experience; let’s call him a “default worker”. Figure 3.1 obviously shows that throughout the decade, the (log) real wage for a default worker has increased significantly at any quantiles. That is, the default worker in 2002 would receive higher pays than if he were in 1995, at the same quantile. This reflects the overall increase in real wages between the years, and is a consequence of the fact that China is becoming richer through the years.

The coefficient of gender is significantly negative at any quantiles in both years, meaning the female group is generally receiving lower wages than the male group. One explanation is that with the same characteristics, female may not be as productive as male, e.g., because females are not as strong as males so that they don't have so much energy as males, leading to relatively lower wages. At the same time, we notice an upward trend of the coefficients in both years, as we move up the conditional wage distribution. This means, within the group of female workers who share the same characteristics, gender bias problem is less severe at higher quantiles. Perhaps high paid jobs have lower physical requirements, while many low paid jobs might have higher physical requirement which put women at disadvantages. This finding shows that wage is more dispersed in female group than in male group (consider two females and twp males who share exactly the same set of characteristics; if one female and one male perform at *0.9th* quantile, while the other two perform at *0.1th* quantile; the difference in real wages between the two females will be $\beta_{gender}^{0.9} - \beta_{gender}^{0.1}$ larger than the difference in real wages between the two males). As we compare 2002 and 1995 gender coefficients in the bottom figure of Figure3.2, we find that the coefficients are much more negative in 2002. This decrease in gender gap indicates a more serious gap between wages of male and wages of female in 2002. This could have contributed toward a larger wage inequality.

It's not surprising to find from Figure3.3 that returns to education are at every quantile significantly positive. A person with more years of education would have higher pays than those with less years of education, with other characteristics the same. An interesting thing here is that as we move up the conditional quantile distribution of wages, the positive effect of education seems to be diminishing (see from the obvious downward sloping trend), implying that high paid jobs pay relatively less to education qualification while low paid jobs pay relatively more. Let's compare the difference of returns to education between the two years. Returns to education are nearly anywhere higher in 2002, compared with 1995. For example, one more year of education would increase a worker's wage by 11.2% in 2002 while it would increase the wage of the same worker by only 6.8% in 1995, at the *25th* conditional quantile of wage distribution (read from Table3.1 and Table3.2). The notable increase in rates of return to education has no doubt increased the overall wage level for the educated, and hence contributed to the rightward shift of the wage density curve. Besides, the overall change in rates of return to education is sloping upward, indicating a larger increase of wages at high quantiles than that at low quantiles from 1995 to 2002. For example, from Table3, rate of return to education increased from 8.7% in 1995 to 11.5% in 2002 at *10th* quantile while it increased from 4.1% in 1995 to 8.0% in 2002 at *90th* quantile. That is, high paid jobs' payoffs for education qualification have increased more than that of low paid jobs, resulting in an increased wage inequality.

Years of potential experiences appear in our model with both linear and quadratic terms. Although Figure3.4 and 3.5 have provided the coefficient estimates of linear term and quadratic term respectively, it's difficult to see the overall effect of potential experience from the two figures. If we look at Figure 3.4 alone here, the estimates of the returns even seem counter-intuitive, as one would expect high paying jobs to reward experience more than low paying jobs; this might because of the inclusion of quadratic potential experience in the model. Hence, in order to see the effect of potential experience directly, we evaluated the overall effect of potential experience at the mean potential experience value, plotted in Figure3.6. That is, we plotted $\beta_{p\exp} + \beta_{p\exp_s} * mean(p\exp)$, since this is the first order derivative of potential experience evaluated at the mean value in our model. It's true for both years that workers with more potential experiences will have higher pays, especially so at low paid jobs. But the change of effect between the years is negligible, indicating no contribution of potential experience toward increased wage inequality.

4.2 Counterfactual Decomposition Analysis

Following the procedures described in Section2.3, we have estimated the marginal wage* densities, denoted as $f(y*(2002);x(2002))$ for 2002 and $f(y*(1995);x(1995))$ for 1995, that are consistent with the linear conditional quantile regression model (*) in Section2.2. Furthermore, we have also estimated the 2002 counterfactual wage* density $f(y*(2002);x(1995))$ and 1995 counterfactual wage* density $f(y*(1995);x(2002))$.

Figure4 performs the specification test of our model. It plots the comparisons of the actual marginal wage density directly obtained from the data with the estimated marginal wage* density that is consistent with our quantile model (*), for both years. The left panel shows that the estimated wage* for 1995 (blue curve) has more or less the same density as the actual 1995 wage distribution (dotted curve). Similarly, in the right panel, our estimated wage* for 2002 (red curve) resembles the actual 2002 wage distribution (dotted curve). Obviously, our quantile regression model is doing a pretty accurate job in estimating the wage distributions.

In Figure5, we look at the impact of coefficients, i.e. the difference between wage densities that are obtained with the same set of covariates but different sets of coefficients. The left panel in the first row plots 1995 and 2002 wage* densities. Consistent with the true density plots in Figure1, 2002 wage* density (red curve) is more to the right and shorter than 1995 wage* density (blue curve), which indicates an increased overall wage level and increased wage inequality. The right panel is a legend box, listing the legends of different line colors. In row two, we compare the 2002 counterfactual wage* density (purple curve) with 1995 estimated wage* density (blue curve) in the left panel. Indisputably, the difference between 2002 counterfactual wage* density and 1995 wage* density is almost the same with the difference between 2002 and 1995 marginal wage* densities. That is, even if the work force in 2002 were distributed as in 1995, because of the change in rates of return, wage densities between the two years would still be different in the same manner. Meanwhile, we also provide the reverse order plots – comparison between 2002 wage* density and 1995 counterfactual wage* density (green curve). Obviously, even if 1995 work force were distributed as in 2002, the difference of wage densities between the two years would be in the same manner as the true difference. Therefore, we have evidence that the change in rates of return to characteristics between the two years contributes significantly toward the change in wage distributions.

In contrast, the impact of covariates (difference between density curves that share the same set of coefficients but different sets of covariates) is surprising. It’s apparent from the left panel in Figure6 row two that there is a slight difference between 2002 wage* density and 2002 counterfactual wage* density (purple curve) – 2002 counterfactual wage* density is a bit shorter than 2002 wage* density; this implies that if the work force characteristics in 2002 were exactly like that of 1995, then the wage inequality should have been more serious. The reverse order comparison in the right panel confirms that if the work force in 1995 were as in 2002 (indicated as green curve), then 1995 wage inequality would be less serious. Therefore, quite opposite to the findings of other countries, like USA and Germany (where change in work force characteristics have enlarged wage inequality), in urban Chinese economy, the change in work force characteristics have contributed toward decreased wage inequality.

5. Conclusion

The phenomenon of increased wage inequality has been observed in many countries during the last decade. In this paper, we analyzed the change in earning distributions of urban Chinese economy between the year 1995 and 2002. The datasets used in our paper are obtained from the China Household Income Project, 1995 and 2002. When we examined the two datasets, we found that the composition of the work force has changes during the period. The annually real wage of workers (full-time employees with ages between 18-55 years old) from urban areas of China has increased from 1995 to 2002. This is mainly due to the fact that China is becoming much richer. Besides, 2002 wage density displays larger wage inequality than 1995 wage density.

To estimate the wage densities in both years, we applied linear quantile regression on the conventional Mincerian wage-education equation on each year's dataset respectively. Our specification test proves that our models are highly accurate. We found that gender gap has decreased significantly while rate of return to education has increased significantly between the two years. In order to know what exactly have caused the change in wage distributions, we have done a counterfactual decomposition analysis with the help of the methodology used in *J.A.F. Machado and J. Mata's paper (2005)*; we found that changes in work force characteristics have contributed slightly toward decreased wage inequality while changes in rates of return to work force characteristics, especially changes in return to education and changes in gender bias, have contributed significantly toward the increased wage inequality.

Nevertheless, there are still areas which need further research. For example, some of the increasing inequality in wages may have been caused by the increasing share of migrant workers in the workforce, which will not be discussed here.

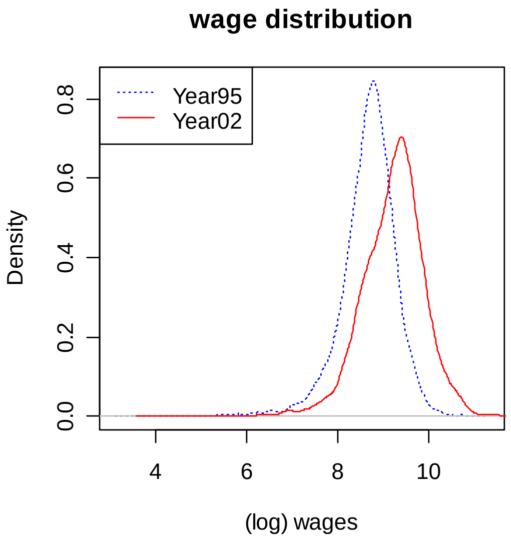
Reference

1. Roger Koenker and Kevin F. Hallock (2001), “Quantile Regression”. Journal of Economic Perspectives, Vol.15, No.4, pp143-156.
2. Roger Koenker (2005), “Quantile Regression”. Cambridge University Press.
3. James L. Powell (1986), “Censored Regression Quantiles”. Journal of Econometrics, Vol.32, pp143-155.
4. Moshe Buchinsky and Jinyong Hahn (1998), “An Alternative Estimation for the Censored Quantile Regression Model”. Econometrica, Vol.66, No.3, pp653-671.
5. Tae-Hwan Kim and Halbert White (2002), “Estimation, Inference, and Specification Testing for Possibly Misspecified Quantile Regression”. Advances in Econometrics, Vol.17, pp107-132.
6. Nils Lid Hjort and David Pollard (1993), “Asymptotics for minimisers of convex processes”. University of Oslo and Yale University.
7. Joshua Angrist, Victor Chernozhukov and Ivan Fernandez-Val (2006), “Quantile Regression under Misspecification, With an Application to the U.S. Wage Structure”. Econometrica, Vol.74, No.2, pp539-563.
8. Yeo Khee Yong, Th Mun Heng, Shandre Mugan Thangavelu and James Wong (2007), “Premium on Fields of Study: The Returns to Higher Education in Singapore”.
9. Haoming Liu (2010), “Human Capital Incestments and Gender Earnings Gap: Evidence from China’s Economic Reforms”.
10. Jose A.F. Machado and Jose Mata (2005), “Counterfactual Decomposition of Changes in Wage Distributions Using Quantile Regression”. Journal of Applied Econometrics, Vol.20, pp445-465.
11. Enrico Moretti (2010), “Real Wage Inequality”.

12. Christopher H. Wheeler (2005), “Evidence on Wage Inequality, Worker Education, and Technology”. Federal Reserve Bank of St. Louis Review, May/June 2005, 87(3), pp. 375-93.
13. John A.Bishop, Feijun Luo and Fang Wang (2004), “Economic Transition, Gander Bias, and the Distribution of Earnings in China”.
14. Chuliang Luo (2008), “The Return to Education and Its Distribution in Urban China: Evidence from Quantile Regression Analysis”. China Economic Journal, Vol.1, pp165-175.
15. Roger Koenker, “Quantile Regression in R: A Vignette”.
16. Alberto Abadie (1997), “Changes in Spanish Labor Income Structure During the 1980's: A Quantile Regression Approach”. Investigaciones Economicas, Vol. XXI(2), pp253-272.
17. John S. Heywood (1988), “The Union Wage Profile of Women: Potential vs. Actual Experience”. Economics Letter, Vol.27, pp189-193.
18. Miller, Carole F. (1993), “Actual Experience, Potential Experience or Age, and Labor Force Participation by Married Women”. Atlantic Economic Journal, Dec 1st, 1993.

Appendix

Figure 1



Unconditional density curves for (log) real wages; blue dotted curve represents 1995 distribution and red curve represents 2002 distribution.

Table 1

	Min.	1st Qu.	Median	3rd Qu.	Max	Mean	SD
1995	3.35	8.37	8.73	9.05	11.09	8.67	0.612
2002	3.87	8.81	9.27	9.63	11.98	9.21	0.670

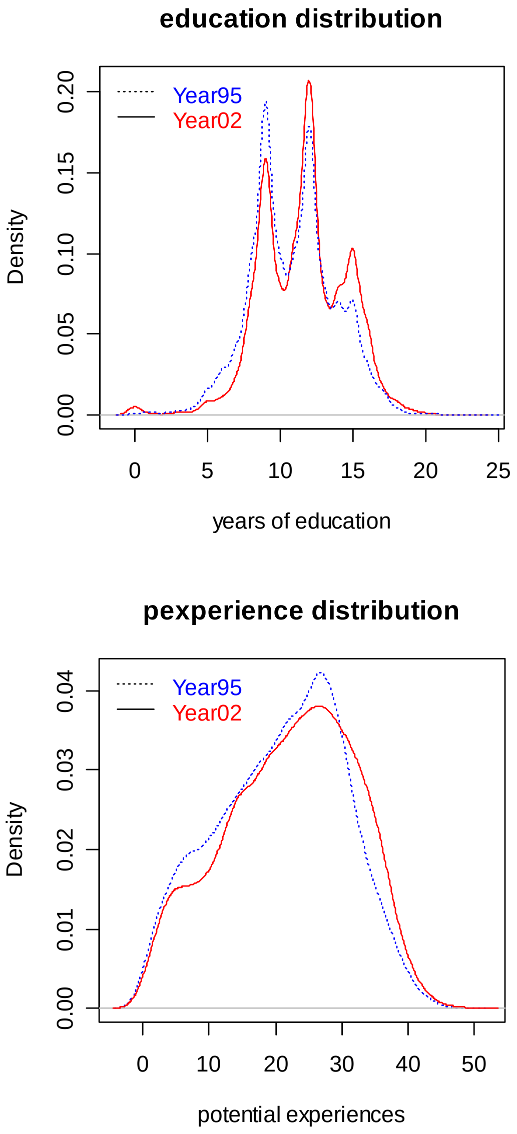
Summary statistics for (log) real wages of both years

Table 2

	1995	**2002**
Observations	8180	7347
Gender (% of female)	48.4	45.5
Years of Education	10.9 (2.9)	11.6 (2.9)
Years of Education (%)		
0	0.1	0.6
6	5.2	2.4
9	32.5	25.8
12	36.2	38.6
16	23.8	29.6
24	2.2	3.0
P-Experience	20.8 (9.5)	22.2(9.7)

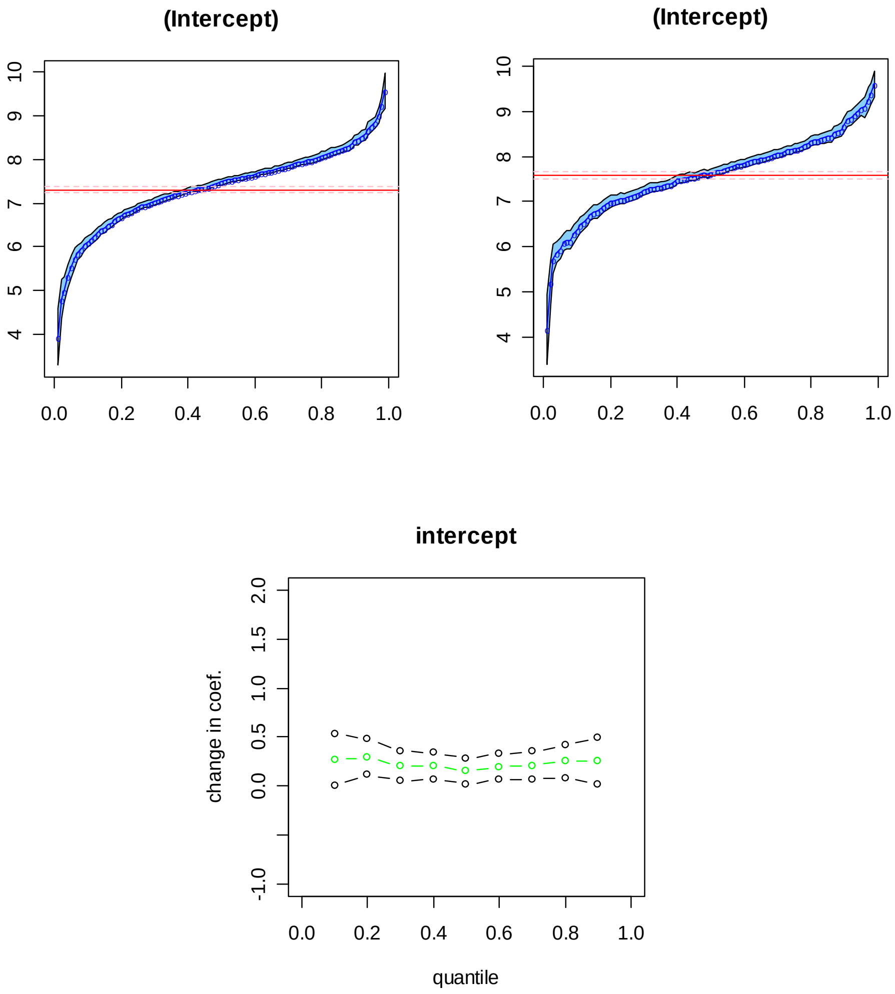
Summary of variables descriptions, with sample standard errors in brackets when sample mean is provided.

Figure 2



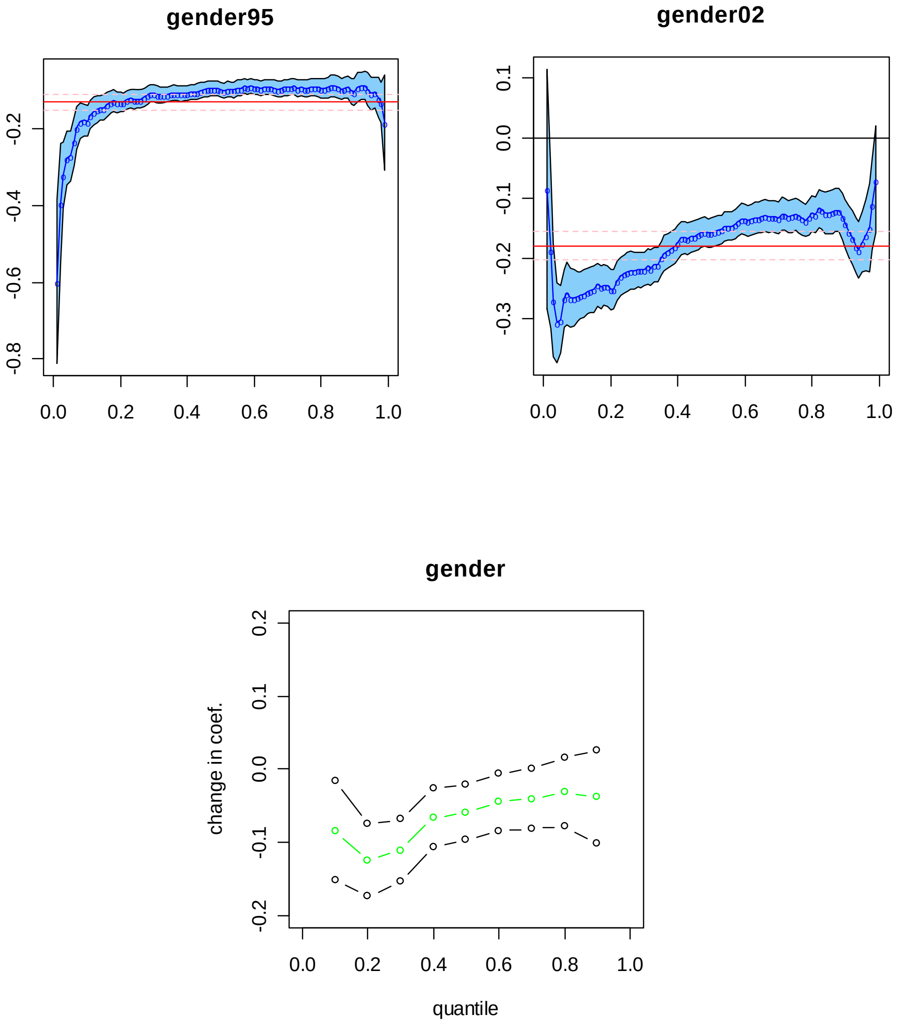
Continuous variables’ distributions; blue dotted curves represent 1995 and red curves represent 2002

Figure 3.1



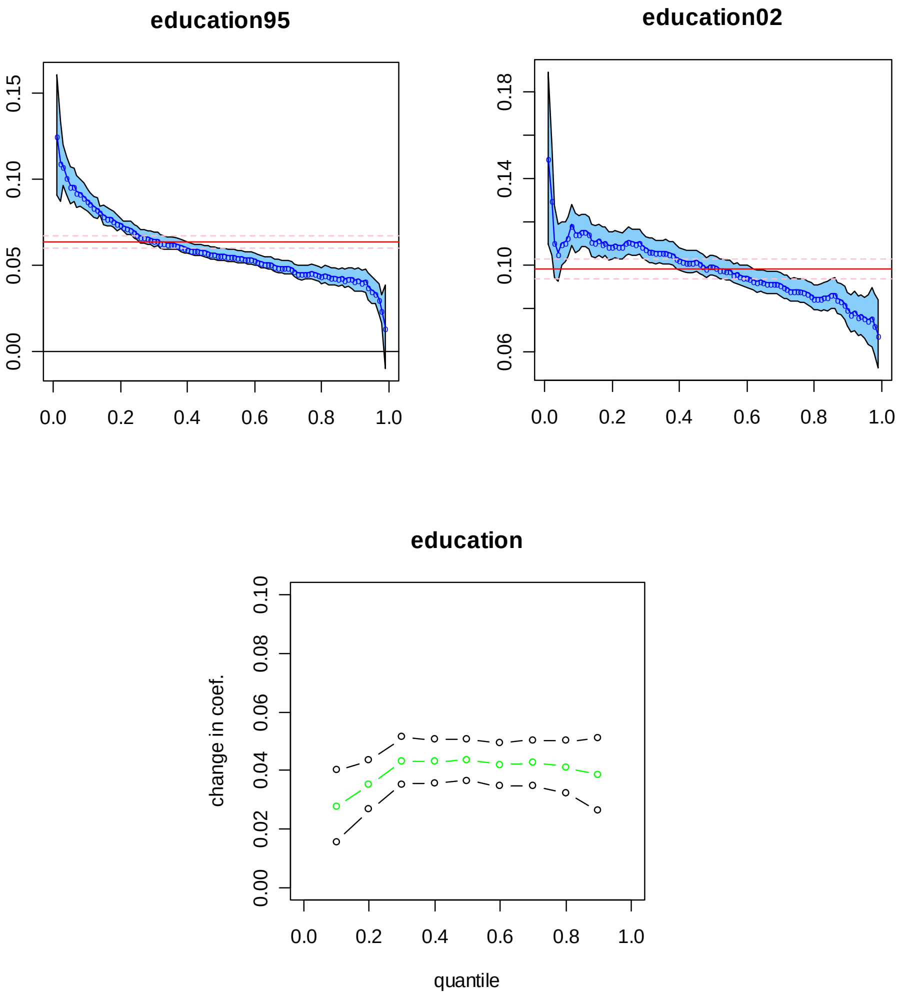
Estimation of intercepts (Top-left figure for estimates of 1995, top-right figure for estimates of 2002, and bottom figure for the change in intercept between the two years)

Figure 3.2



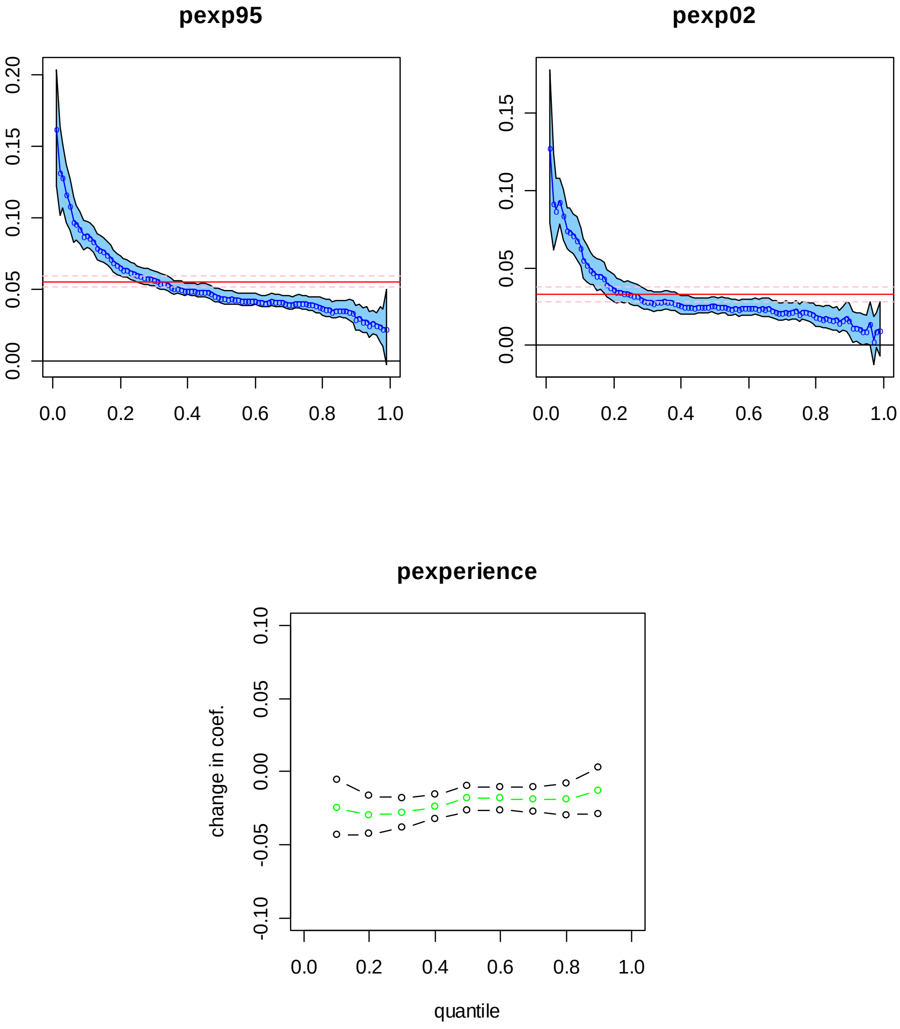
Estimation of gender coefficients (Top-left figure for estimates of 1995, top-right figure for estimates of 2002, and bottom figure for the change in gender coefficients between the two years)

Figure3.3



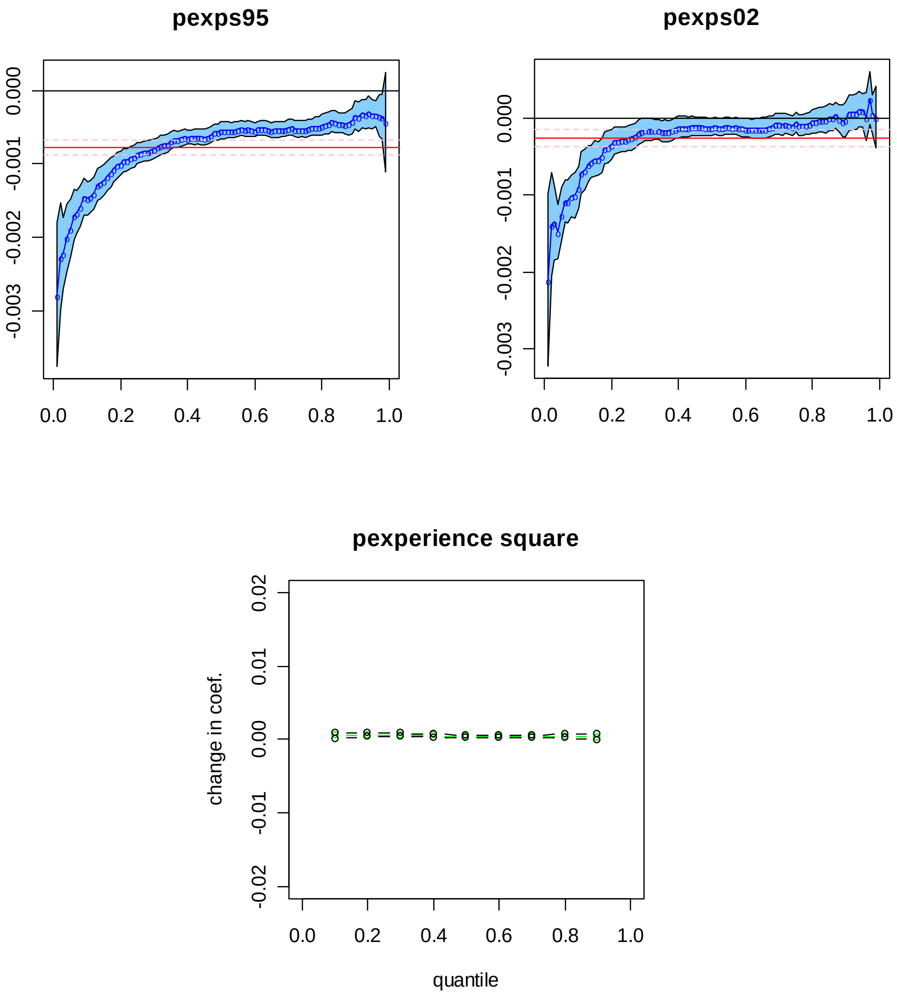
Estimation of education coefficients (Top-left figure for estimates of 1995, top-right figure for estimates of 2002, and bottom figure for the change in education coefficients between the two years)

Figure 3.4



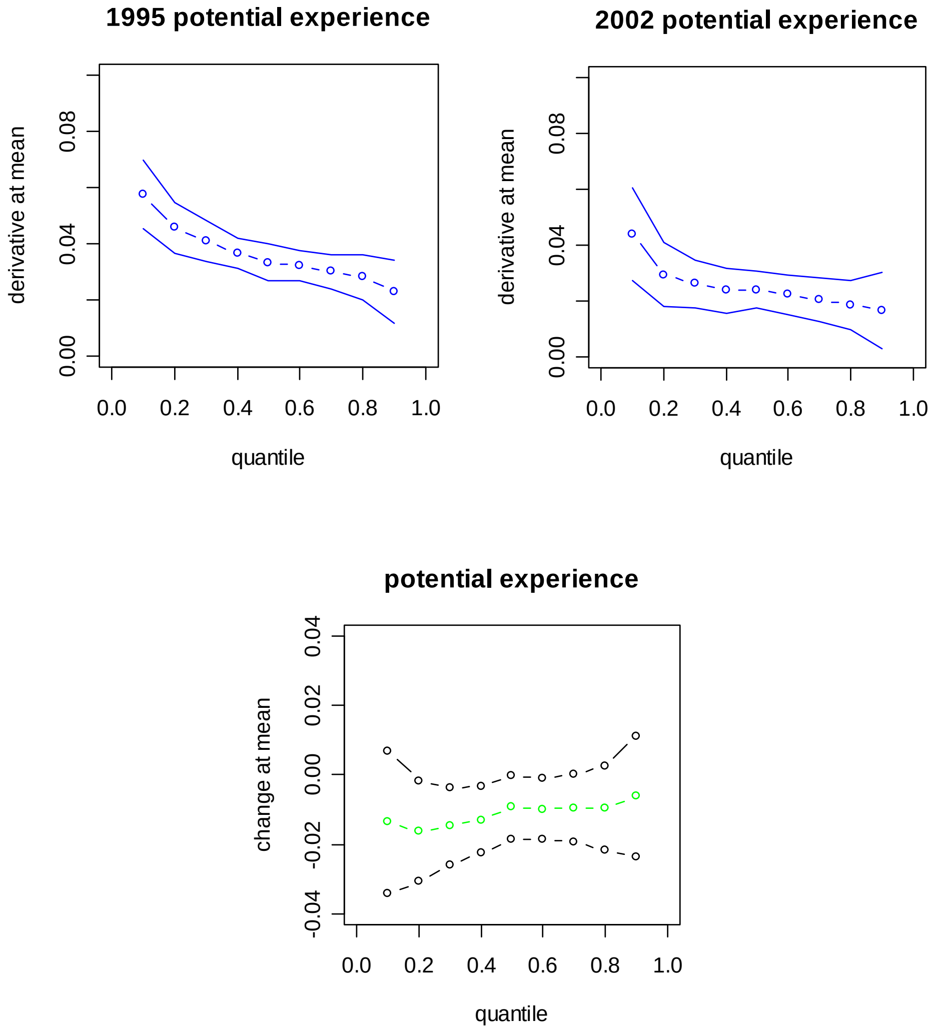
Estimation of potential experience coefficients (Top-left figure for estimates of 1995, top-right figure for estimates of 2002, and bottom figure for the change in potential experience coefficients between the two years)

Figure 3.5



Estimation of potential experience square coefficients (Top-left figure for estimates of 1995, top-right figure for estimates of 2002, and bottom figure for the change in potential experience square coefficients between the two years)

Figure 3.6



Effect of potential experience evaluated at the mean value of potential experience (Top-left figure for estimates of 1995, top-right figure for estimates of 2002, and bottom figure for the change in overall effect between the two years)

Table 3.1

1995	τ =0.1	τ =0.25	τ =0.5	τ =0.75	τ =0.9	OLS
(Intercept)	6.127*	6.902*	7.485*	7.957*	8.435*	7.303*
	(0.077)	(0.048)	(0.042)	(0.044)	(0.073)	(0.039)
Gender	-0.179*	-0.121*	-0.097*	-0.093*	-0.103*	-0.130*
	(0.024)	(0.014)	(0.012)	(0.014)	(0.021)	(0.012)
Education	0.087*	0.068*	0.056*	0.046*	0.041*	0.063*
	(0.004)	(0.002)	(0.002)	(0.002)	(0.004)	(0.002)
P-Experience	0.088*	0.061*	0.045*	0.045*	0.030*	0.055*
	(0.006)	(0.003)	(0.003)	(0.003)	(0.0057)	(0.003)
P-Experience^2	-0.001*	-0.001*	-0.001*	-0.001*	-0.0003*	-0.0008*
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)

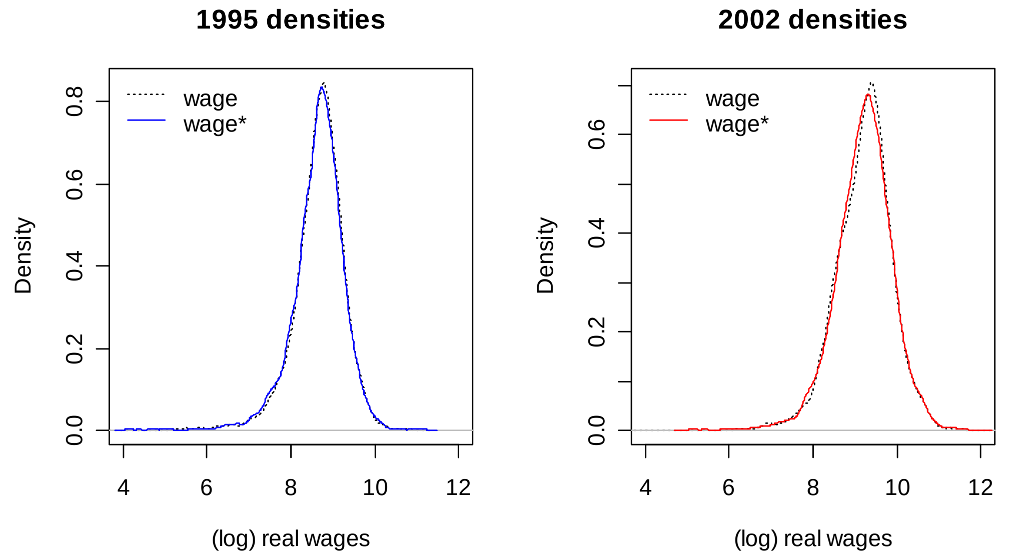
OLS and some typical quantile regression estimates of coefficients, where * indicate significant at 5% significance level and the number in each bracket shows the corresponding sample standard error.

Table 3.2

2002	τ =0.1	τ =0.25	τ =0.5	τ =0.75	τ =0.9	OLS
(Intercept)	6.393*	7.083*	7.637*	8.181*	8.696*	7.585*
	(0.112)	(0.069)	(0.052)	(0.058)	(0.097)	(0.050)
Gender	-0.262*	-0.222*	-0.157*	-0.127*	-0.141*	-0.179*
	(0.025)	(0.020)	(0.015)	(0.016)	(0.025)	(0.014)
Education	0.115*	0.112*	0.100*	0.089*	0.080*	0.099*
	(0.005)	(0.004)	(0.003)	(0.003)	(0.005)	(0.003)
P-Experience	0.064*	0.034*	0.026*	0.021*	0.017*	-0.033*
	(0.008)	(0.004)	(0.003)	(0.003)	(0.006)	(0.003)
P-Experience^2	-0.001*	-0.0003*	-0.000	-0.000	-0.000	-0.0003*
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)

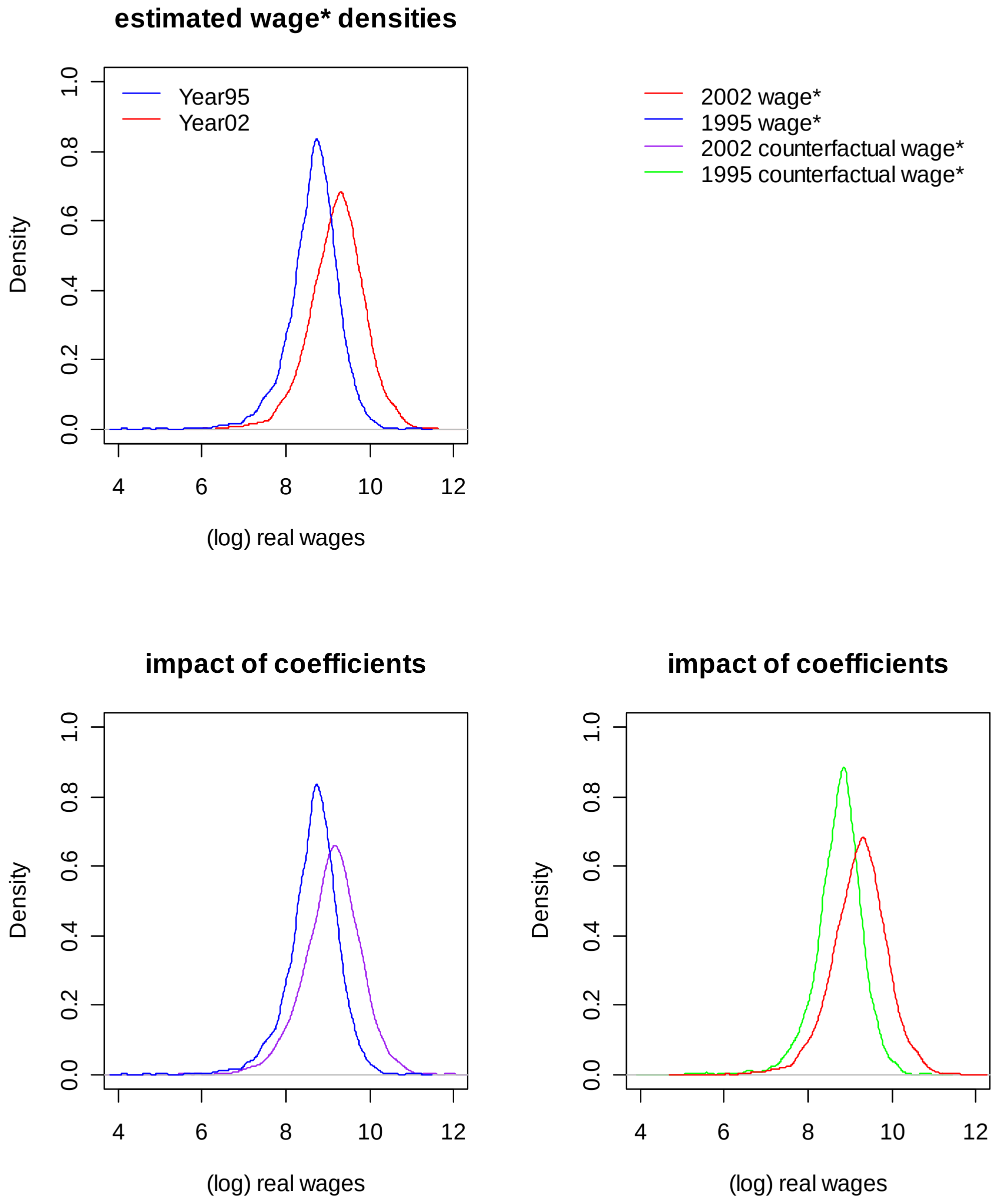
OLS and some typical quantile regression estimates of coefficients, where * indicate significant at 5% significance level and the number in each bracket shows the corresponding sample standard error.

Figure 4



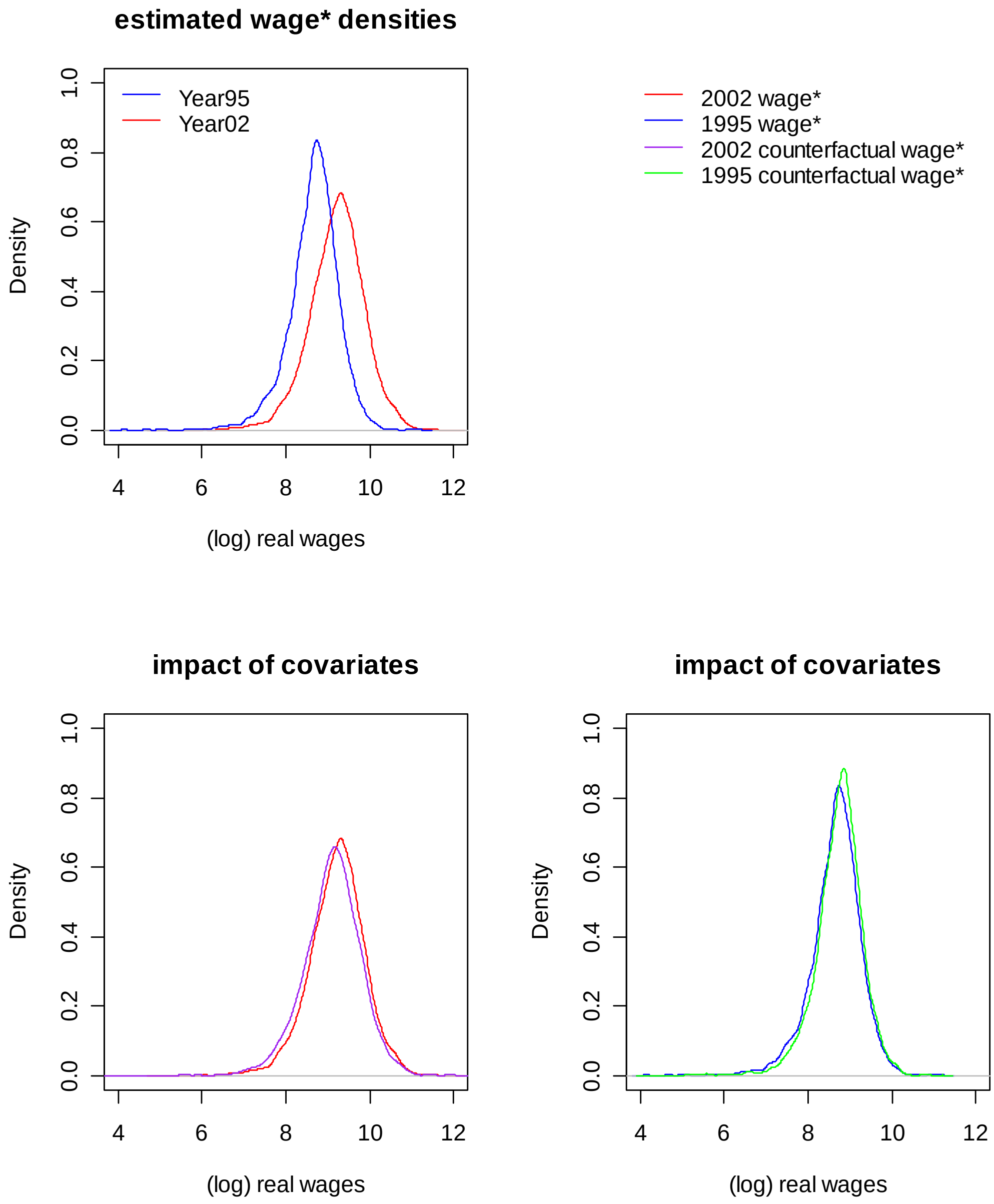
Comparison of actual marginal wage density with the estimated marginal wage* density; left panel for 1995 (blue curve represents wage* density and dotted curve for actual density) and right panel for 2002 (red curve represents wage* density and dotted curve for actual density)

Figure 5



Impact of coefficient changes toward wage distribution

Figure 6



Impact of covariate changes toward wage distribution