DISCUSSION PAPER SERIES

IZA DP No. 3963

Changes in Wage Structure in Urban India 1983-2004: A Quantile Regression Decomposition

Mehtabul Azam

January 2009

Forschungsinstitut
zur Zukunft der Arbeit
Institute for the Study
of Labor

Changes in Wage Structure in Urban India 1983-2004: A Quantile Regression Decomposition

Mehtabul Azam
Southern Methodist University
and IZA

Discussion Paper No. 3963
January 2009

IZA

P.O. Box 7240
53072 Bonn
Germany

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

IZA Discussion Paper No. 3963
January 2009

ABSTRACT

Changes in Wage Structure in Urban India 1983-2004: A Quantile Regression Decomposition*

This paper examines changes in the wage structure in urban India during the past two decades (1983-2004) across the entire wage distribution using the Machado and Mata (2005) decomposition approach. Real wages increased throughout the wage distribution during 1983-1993; however, it increased only in the upper half of the wage distribution during 1993-2004. Quantile regression analysis reveals that the effects of many covariates are not constant across the wage distribution. Moreover, increases in returns to covariates across the entire distribution are the driving forces behind the wage changes in both decades. Change in composition of the work force contributed positively to wage growth during 1983-1993, but negatively during 1993-2004. Finally, while workers with all education levels experienced an increase in returns of roughly the same magnitude during 1983-1993, the increase in returns is much higher for workers with tertiary and secondary education during 1993-2004. The inequality increasing effects of tertiary education suggests that wage inequality in urban India may increase further in the near future as more workers get tertiary education.

JEL Classification: J30, J31, C15

Keywords: earning functions, India, quantile regression decomposition, wage

Corresponding author:

Mehtabul Azam
Department of Economics
Southern Methodist University
Dallas, TX 75275
USA
E-mail: mazam@smu.edu

* I thank Daniel Millimet, Esfandiar Maasoumi, Thomas Fomby and Thomas Osang for helpful comments. Any remaining errors remain mine.

1 Introduction

In this paper, we examine changes in the wage structure in urban India across the entire wage distribution over the past two decades (1983-2004). We investigate the earnings function at three points in time (1983, 1993, and 2004) using ordinary least squares and quantile regressions to assess whether the entire earnings distribution is affected uniformly by human capital variables, demographic characteristics, and industry affiliations. We also investigate the changes in returns to various characteristics over 1983-1993 and 1993-2004. As we describe below, important changes took place over both time periods; however, the institutional settings differ significantly between these two time periods. Furthermore, we apply a quantile regression based decomposition method proposed in Machado and Mata (2005) to evaluate the role of changing labor force composition (in terms of workers' characteristics) and changing labor market prices in overall changes in the wage distribution over these two time periods.

The past two decades are interesting as two important changes took place in India during this period. First, in the 1980s there was a dramatic increase in the average growth rate of GDP to around 5.7 percent per year compared to the 'Hindu rate of growth' of around 3.5 percent that prevailed during 1950-1980. Second, economic liberalization in 1991 abolished the four decade old import-substitution industrialization strategy, and initiated a drastic liberalization of the external sector and industrial policy. The number of industries reserved for the public sector was halved.[1] Licensing was abolished in all but a small group of industries. Foreign investment was encouraged, quantitative import restrictions were largely abandoned and tariffs were significantly reduced. After these changes, the Indian economy grew at an average growth rate of 5.9 percent between 1992-93 and 2002-03 (Panagariya, 2004). Although from a pure growth perspective the 1980s and 1990s were not very different, the institutional settings in these two decades differ significantly. India was very much a

[1] The number of industries reserved for the public sector went down from 17 in 1956 to 8 in 1991 (Economic Survey, 1991-92).

closed economy with a large number of industries reserved for the public sector during the 1980s, whereas the 1990s was a period of integration with the world economy with a growing role of private participation and foreign investment.[2]

However, the economic growth of the past two decades has been associated with rising wage inequality, and this increase in wage inequality is distributed over both decades (Kijima, 2006, Dutta, 2005). Dutta (2005) examines trends in wage dispersion during 1983-1999 using various indices and a regression-based decomposition, and finds that wage inequality increased during 1983-1999. Kijima (2006) examines changes in wage inequality in urban India during 1983-1999 using the Juhn et al. (1993) method, and finds that wage inequality in urban India started increasing before 1991. He attributes the increase in wage inequality over 1983-1999 to an increase in the returns to tertiary education. Chamarbagwala (2006) uses the demand and supply framework of Katz and Murphy (1992) and finds that the skill premium in India increased during 1983-1999, and attributes this increase to skill biased technological change.

A shortcoming of the existing literature on wages in India is that it primarily concentrates on averages, neglecting the rest of the distribution.[3] However, averages may miss important features of the wage structure, and it is important to go beyond averages to present a complete picture for three reasons. First, recent work in other countries using quantile regression techniques have shown that attributes have different effects on the wages of individuals at the top of the wage distribution compared with individuals at the bottom of the wage distribution.[4] Second, India is a heterogeneous society in the midst of rapid

[2] Some policy reforms were introduced during 1980s also. Panagariya (2004) points out that the difference between the reforms in the 1980s and those in the 1990s is that the former were limited in scope and without a clear roadmap, whereas the latter were systematic. Joshi and Little (1994, chapter 13) also recognize the role of the reforms but regard fiscal expansion financed by external and internal borrowing as the key to the acceleration of growth during the 1980s. This is also the view expressed indirectly by Ahluwalia (2002) who states that while the growth record in the 1990s was only slightly better than in the 1980s, the 1980s growth was unsustainable, "fuelled by a buildup of external debt that culminated in the crisis of 1991".

[3] An exception is Kijima (2006) which decomposes the changes in the 90^{th}-10^{th}, 90^{th}-50^{th}, and 50^{th}-10^{th} percentile of log wage differential.

[4] The evidence for this comes from a number of different countries such as the USA (Buchinsky, 1994), Germany (Fitzenberg and Kurz, 2003), Uruguay (González and Mile, 2001), Zambia (Nielson and Rosholm, 2001), and Portugal (Machado and Mata, 2000).

change. This suggests that effects may be heterogeneous as well. Third, there is growing evidence from other countries (e.g., the US) that suggest that, far from being ubiquitous, the growth in wage inequality is increasingly concentrated in the top end of the wage distribution (Lemieux, 2007).

This paper contributes to the existing literature in the following way. First, we estimate earning functions across the entire wage distribution using quantile regression, and analyze the changes in the contribution of individual covariates over time. Second, we decompose the change in wages in the past two decades into a part that is attributable to a change in prices (the coefficient effect) and a part that is attributable to a change in characteristics (the covariate effect) across the entire wage distribution. Third, we extend the existing literature through 2004 by incorporating new data.

The findings of the paper are as follows. First, real wages increased across the entire wage distribution between 1983 and 1993; however, it increased only in the upper half of the wage distribution between 1993 and 2004. Second, the wage changes are driven mostly by an increase in prices paid (coefficient effect) across the entire wage distribution in both decades. Also, this coefficient effect is larger at higher quantiles. The change in the characteristics of the work force (covariate effect) contributed positively between 1983 and 1993. However, the covariate effect was negative between 1993 and 2004. Third, returns to many characteristics are not homogeneous in the three years of study. Returns to tertiary education not only increased by almost 20 percent between 1993 and 2004, but also became more heterogeneous in 2004. Fourth, decomposing the wage changes for different education groups separately, we find that while workers with all education levels experienced approximately similar increases in prices paid (positive coefficient effect) between 1983 and 1993, the increase in prices paid are much higher for workers with tertiary and secondary education between 1993 and 2004. Also, the positive coefficient effect for workers with tertiary and secondary education is highly heterogeneous across quantiles.

The findings of the paper suggest that wage inequality in urban India will continue to increase in the near future as educational composition changes over time and more workers obtain tertiary education. Since returns to tertiary education have not only increased but also have become more heterogeneous over the last decade, it will add to both within and between group inequality in the near future. However, the increase in skill premium during the 1990s is expected to stimulate further increases in human capital investment (Topel, 1999), and an increase in the number of college graduates may decrease the wage inequality in the long run as South Korea experienced in 1970s and 1980s (Kim and Topel, 1995).

Also, there are a large number of developing countries which, like India, chose a strategy of import-substitution as a mean of industrializing after the Second World War. In the past two decades, many of them have begun to favor global economic integration, and in particular trade liberalization, as a development strategy. A number of developing countries that opened up to trade more recently had similar experiences like India. Liberalization did not lead to a reduction in wage inequality; on the contrary, it has increased both wage inequality and the skill premium (Arbache et al., 2004).[5] Although the effects of trade liberalization on wages are extensively studied in most of these countries, most of the studies concentrate on averages only. However, from the welfare perspective it is important to go beyond the mean and see how the whole distribution of wages has evolved.

The remainder of the paper is organized as follows. Section 2 deals with the empirical strategy, Section 3 describes the data, Section 4 investigates the results, and Section 5 concludes.

[5] Arbache et al. (2004) survey increases in the wage premium following liberalization in Brazil (Green et al., 2001), Mexico (Hanson and Harrison, 1999; Robertson, 2000), Chile (Beyer, Rojas, and Vergara, 1999), Morocco (Currie and Harrison, 1997), Costa Rica (Robbins and Gindling, 1999) and Columbia (Robbins, 1996a).

2 Empirical Strategy

2.1 Quantile Regression

Let $Q_\theta(w|x)$ for $\theta \in (0,1)$ denote the θ^{th} quantile of the distribution of the log wage given the vector of covariates x. We model these conditional quantiles as:

$$Q_\theta(w|x) = x'\beta(\theta) \tag{1}$$

where $\beta(\theta)$ is a vector of quantile regression (QR) coefficients (Koenker and Basset).

2.2 Decomposition of Changes in the Wages

2.2.1 Standard Decomposition

Before discussing how to decompose the wage changes across the entire wage distribution, it is useful to discuss the familiar case of decomposing differences in the mean where the standard Oaxaca-Blinder decomposition can easily be used. Consider a standard (log) wage equation for year t:

$$w_t = x_t\beta_t + \varepsilon_t, \quad t = 0, 1 \tag{2}$$

Under the usual assumption that the error term ε_t has a conditional mean of zero given the covariates x_t, β_1 and β_0 can be consistently estimated using OLS and the mean difference in wages between year 1 and year 0 can be decomposed as:

$$\Delta = \overline{w_1} - \overline{w_0} = \underbrace{\overline{x_1}(\beta_1 - \beta_0)}_{\text{Coefficient effect}} + \underbrace{(\overline{x_1} - \overline{x_0})\beta_0}_{\text{Covariate effect}} \tag{3}$$

where $\overline{w_1}$ and $\overline{w_0}$ are the mean wages in year 1 and year 0, respectively, while $\overline{x_1}$ and $\overline{x_2}$ are the corresponding mean values of explanatory variables.

2.2.2 Decomposition of Differences in Wage Distributions

We would like to perform a similar decomposition across the entire wage distribution. A number of decomposition procedures have been suggested to untangle the sources of differences between wage distributions. Popular methods used in the wage inequality literature include the "plug-in" procedure of Juhn, Murphy, and Pierce (1993, JMP hereafter) based on parametric regressions, the reweighting procedure of DiNardo, Fortin, and Lemieux (1996), and more recently a quantile regression based decomposition method of Machado and Mata (2005). The limitation of the JMP (1993) methodology is that it is valid only in the case of homoskedasticity, which is usually rejected for empirical wage distributions. In addition, it is restrictive as it assumes a single linear model to hold for the entire distribution. The limitation of the DiNardo, Fortin, and Lemieux (1996) technique is that it allows one to investigate the role of changes in endowments only.[6] In addition, it relies on OLS regressions to do quantile analysis.

In this paper, we apply the Machado and Mata (2005, MM hereafter) technique. The advantage of the MM technique is that the quantile regressions account for heteroskedasticity, and it partitions the observed difference in wage distributions into 'price' and 'quantity' components. The MM (2005) decomposition is well suited to depict heterogeneous characteristic and coefficient effects across the entire distribution. As demonstrated by Autor et al. (2005), the MM approach nests most of the usual approaches.

The first idea underlying the MM technique is that the conditional quantiles of w, given by equation (1), can be estimated by quantile regression. If equation (1) is correctly specified, the conditional quantile process – that is, $Q_\theta(w|x)$ as a function of $\theta \in (0, 1)$ – provides a full characterization of the conditional distribution of wages given x. In more general cases, the conditional quantile model may provide a reasonable approximation to the true conditional quantile.

[6] Leibbrandt et al. (2005) proposes a simple extension to DiNardo, Fortin and Lemieux (1996) to separate the role of quantities and prices in influencing the shape of the wage distribution.

The second idea underlying the MM technique is the probability integral transformation theorem: If θ is a uniform random variable on $[0,1]$, then $F^{-1}(\theta)$ has the distribution F. Thus, if $\theta_1, \theta_2, ..., \theta_m$ are drawn from a uniform $(0,1)$ distribution, the corresponding m estimates of the conditional quantiles of wages at x_i, $\widehat{w}_i = \{x_i'\widehat{\beta}(\theta)\}_{i=1}^m$, constitute a random sample from the (estimated) conditional distribution of wages. To 'integrate x out' and get a sample from the marginal distribution, instead of keeping x fixed at a given value, a random sample can be drawn from appropriate distribution. The algorithm below summarizes the procedure:

Let $w(t), x(t), t = 1983, 1993, 2004$, denote wages and the k covariates in year t. In addition, let $g(x;t)$ be the joint density of the covariates in year t. We want to generate a random sample from the wage density that would prevail for year t if the conditional model (1) were true and the covariates were distributed as $g(x;t)$. To get this,

1. Generate a random sample of size m from a uniform $(0,1)$: $\theta_1, \theta_2, ..., \theta_m$.

2. For the data set for year t (denoted by $x(t)$, a $n_t \times k$ matrix of data on the covariates), and each $\{\theta_i\}$ estimate

$$Q_{\theta_i}(w|x;t)$$

which yields m estimates of the QR coefficients $\widehat{\beta}(\theta_i)$.

3. Generate a random sample of size m with replacement from the rows of the covariate matrix $x(t) : \{x_i^*(t), i = 1, 2, 3....m\}$.

4. Finally

$$\{w_i^*(t) = x_i^*(t)'\widehat{\beta}(\theta_i)\}_{i=1}^m$$

is a random sample of size m from the desired distribution.

However, instead of drawing a random sample of size m from uniform distribution and estimating $\widehat{\beta}(\theta)$ for each θ_i, we estimated $\widehat{\beta}(\theta)$ on a grid of $\theta's = [0.001, 0.002, ..., 0.998, 0.999]$. Then we draw $m = 1000$ random draws from the distribution of covariates for each $\widehat{\beta}$ and stack $x'\widehat{\beta}$ to get the desired distribution.[7],[8] In practice, this procedure yields the same estimate and removes the sampling error from the first step of MM procedure (Albrecht et al., 2007).

Alternative distributions could be estimated by drawing x from another distribution and using different coefficient vectors. As noted by Autor et al. (2005), this procedure is equivalent to numerically integrating the estimated conditional function over the distribution x and θ. Suppose we want to estimate the density function of wages in year 1, corresponding to the year 0 distribution of the covariates. To do this we follow the algorithm above for year 1, but in the third step, instead of drawing the sample from $g(x;1)$, we sample it from the rows of $g(x;0)$. After obtaining the desired counterfactual densities we can decompose the overall wage difference between two years into a part attributable to the coefficients, another to the covariates, and a residual.

Let $f(w(1))$ denote the estimate of the marginal density of w (log of wages) in year 1 based on observed sample, i.e., the empirical density, and $f^*(w(1))$ denote an estimate of the density of w in year 1 based on the generated sample $\{w_i^*(1)\}_{i=1}^m$, i.e., the marginal implied by the model. Extending this notation to the counterfactual distributions, we may define $f^*(w(1);x(0))$ as the density that would have prevailed in year 1 if all covariates would have been distributed as in year 0, but the workers are paid as in year 1.

We may use the counterfactual distribution, $f^*(w(1);x(0))$, to decompose the changes in wage distributions between any two years.[9] Letting α be a usual summary statistic (for

[7] To take into account the household survey weights, we implemented unequal probablity sampling with replacement.

[8] We end up with 990,000 observations for w_i^*.

[9] There is another possible counterfactual that can be used in the decomposition, i.e., $f^*(w(0);x(1))$, which is the wage density that would have prevailed if all covariates are distributed as in year 1, but workers are paid as in year 0. It is well known that decomposition results may not be invariant with respect to the choice of the involved counterfactual (Oaxaca and Ransom, 1994). However, qualitatively our results are invariant to the alternative counterfactual, i.e., $f^*(w(0);x(1))$.

instance, quantile or scale measure), we may decompose differences in α as:

$$
\begin{aligned}
& \alpha\{f(w(1))\} - \alpha\{f(w(0))\} \\
= \ & \alpha\{f^*(w(1))\} - \alpha\{f^*(w(0))\} + residual \\
= \ & \alpha\{f^*(w(1))\} + \alpha\{f^*(w(1);x(0))\} - \alpha\{f^*(w(1);x(0))\} - \alpha\{f^*(w(0))\} + residual \\
= \ & \underbrace{\alpha\{f^*(w(1))\} - \alpha\{f^*(w(1);x(0))\}}_{\text{Covariate effect}} + \underbrace{\alpha\{f^*(w(1);x(0))\} - \alpha\{f^*(w(0))\}}_{\text{Coefficient effect}} + residual (4)
\end{aligned}
$$

This decomposition will then give us the contribution of the covariates, the coefficients and an unexplained part (residual).

3 Data

The analysis in this paper draws on individual level data from the Employment and Unemployment Schedule, administered by the National Sample Survey Organization (NSSO), Government of India. We use data from the 38^{th}, 50^{th}, and 61^{st} rounds which were conducted in 1983, 1993-94, and 2004-05, respectively (referred to as 1983, 1993 and 2004 in this paper).[10] The data constitute a repeated cross section and contain information on household size and composition, social group, religion, monthly consumption, landholdings, demographic variables (age, gender, marital status), educational participation and attainment, along with a detailed employment section on principal and subsidiary activities (industry, occupation, type and amount of wages earned) . The sample is drawn based on a stratified random sampling procedure and all the analysis is done using survey weights.

We restrict our attention to the urban labor market.[11] In the data, workers are classified

[10] NSSO conducts thick round surveys (called 'quinquennial rounds') at five-year intervals. The data before 1983 is not available and 2004-05 is the most recent round available. We chose to use 1993-94 data as it divides the past twenty years into two equally spaced time periods corresponding to two different institutional settings. India's decision to liberalize in May 1991 was sudden and externally imposed. One would expect that this sudden policy shift will affect the labor market with significant lag. So 1993-94 serves as a possible benchmark for end of closed economy era and start of an open economy era.

[11] The labor market in rural India is very thin.

as self-employed, regular wage/salaried and casual labor. Weekly wages earned are reported at current prices for regular wage/salaried and casual labor.[12] Reported weekly wages are deflated to 1984-85 prices using the state-specific consumer price index for urban non-manual employees. The spatial differences in cost of living in each year are adjusted using the ratio of the official state-specific urban poverty line to the all India urban poverty line. Since it is difficult to predict wages for casual labor, as their wages may not bear any definite relation with general covariates, we restrict our attention to regular wage/salaried workers.[13] Regular wage/salaried workers are full-time workers with strong labor force attachment. We also examine only those workers who are 21-60 years old.

The 1983 and 1993 data use National Industrial Classification codes (NIC) - 1970 and 1987, respectively, to report industry of employment at three digits, while 2004 data uses NIC-1998 to report industry of employment at five digits. Uniformity is established across the three surveys using a concordance table and 24 broad industries are created.

In this paper, we do not address selection into regular wage/salaried labor market for three reasons. First, the technique required to correct for selectivity bias in quantile regression models is less well developed.[14] Second, even if one could adequately address non-random sample selection, we are only interested in describing the wage distribution conditional on being in regular employment. Third, since we are comparing the same urban labor market over time, and the share of regular workers in the relevant population has not changed much during the time period under study, it is unlikely that a standard correction for sample

[12] A regular wage salaried worker is a person who works in others' farm or non-farm enterprises (household and non-household) and in return received salary or wages on a regular basis; while a casual worker is a person who is engaged in others' farm or non-farm enterprises (household and non-household) and in return, received wages according to the terms of the daily or periodic work contract.

[13] Dutta (2006) finds that while the explanatory power of human capital variables is reasonably high for regular wage/salaried workers, that for casual workers is very low, suggesting that human capital characteristics do not explain the wage determination process for casual labor.

[14] Buchinsky (1998) suggests a way to correct sample selection, and this selection correction is incorporated into the MM technique by Albrecht et al. (2007). However, Buchinsky's selection correction technique requires a valid exclusion variable. Possible exclusion variables available in our data are household structure variables (i.e., number of children, or the dependency ratio, etc.) and landholding. Landholding may be a good exclusion variable in rural areas, but in urban areas it is not very useful. Similarly, household structure variables are suspect since how they affect the employment participation of the predominantly male working population is not very clear.

selection would significantly affect our decomposition results.

Our dependent variable is log of real weekly wage, and the covariate matrix x includes age, age squared, dummies for state, education level, female, married, the Scheduled Castes (SCs), the Scheduled Tribes (STs), Muslims, and industries.[15] The SCs and the STs are two historically disadvantaged groups, and enjoy the benefit of affirmative action taken by Government of India in the form of reservation in jobs and education. Muslims constitute the largest religious minority group in India.[16]

Table 1 presents the descriptive statistics of the sample used, and broad measures of wage inequality (like the standard deviation and the 90-10 gap in log wage). The real wage increased at a rate of 4.1% per annum during 1983-1993, while the rate of increase declined to 3% per annum during 1993-2004. A strong pattern of results emerge from simple comparisons of measures of wage inequality. First, the log wage differential between the 90^{th} and 10^{th} percentile increased by 27 log points during 1983-1993, and by 37 log points during 1993-2004. Second, while both "low-end" (50-10 gap) and "top-end" (90-50 gap) inequality increased (12 and 16 log points, respectively) between 1983 and 1993, the situation changed radically between 1993 and 2004. The 50-10 gap increased by 1 log point, while the 90-50 gap increased by 36 log points. So while the wage inequality (captured by the 90-10 gap) increased in both decades, the increase in wage inequality during 1983-1993 was more evenly distributed across the entire wage distribution whereas rising upper tail inequality (the 90-50 gap) accounts for almost all the increase in the overall expansion in the 90-10 gap during 1993-2004.

Over the entire period (1983-2004), the increase in wage inequality has been mainly due to increase in wages of the groups above the median (the 90-50 gap increased by 48 log points

[15] Illiterates/below primary is treated as base dummy for education levels. Dummies included represent primary, middle, secondary, and tertiary levels of education.

[16] Muslims constitute 13.4 percent of the total population in 2001. A committee constituted by Government of India (popularly known as "Sacher Committee", 2006), to study the "social, economic and educational status of Muslims in India", points out that by and large, Muslims rank somewhat above the SCs/STs but below Hindu other backward castes, Hindu upper castes, or other Minorities in almost all indicators considered.

while the 50-10 gap increased by 17 log points). The findings are consistent with Kijima (2006), who also finds that the increase in wage inequality between 1983 and 1999 has been mainly due to increases in the income of groups above the median. Banerjee and Picketty (2003), using individual tax return data from 1956 to 2000, also find that the income share of the top percentiles increased during the 1980s and 1990s.

Table 2 presents the distribution of workers with different education levels across quintiles in each year. Approximately 48 percent of illiterates/below primary are in the bottom 20 percent of the wage distribution in 2004, while almost 50 percent of tertiary educated workers are in the top 20 percent of the wage distribution. So there appears to be a high correlation between education level achieved and wages earned.

Table 3 presents the shares of different industries in regular wage/salaried employment. Between 1983 and 2004 the employment share of public administration and defense decreased from 23 percent to 14.5 percent, and most of the decrease occurred between 1993 and 2004. This indicates the decreasing importance of the public sector as an employment generator after early 1990s. The share of wholesale and retail trade and computer related activity increased by more than 3.5 percentage points. There are no other major changes in shares of other industries.

Figure 1 plots the kernel density of the log of real weekly wages for three years. Between 1983 and 1993, the whole distribution shifted to the right; however, between 1993 and 2004 the right part of the distribution shifted to the right while the left part remained virtually unchanged from 1993. Figure 2 plots the cumulative distribution function (CDF) of the log of real weekly wages in the same three years. The CDFs cross each other at the initial values of log real wages. So, first order stochastic dominance is not accepted over the entire range of log wages. However, the wage distribution in 2004 first order dominates over the wage distribution in 1983 for all practical values of wages. There is no first order dominance between the 1993 and 1983 wage distributions or between the 2004 and 1993 wage distributions.

4 Results

4.1 Quantile Regression

Figure 3a and 3b plot the coefficient estimates, $\beta_i(\theta)$ for $\theta \in (0,1)$, for education levels and demographic characteristics, respectively. The associated 95% confidence bands are represented by the dots. For each variable, the plots provide information on the coefficient estimates for 1983 (Panel A), 1993 (Panel B), and 2004 (Panel C). For comparison purposes, the coefficients estimated by mean regression (OLS) are reported as a dashed horizontal line. Figure 4a and 4b plot the change in estimated coefficient over time at each quantile for education levels and demographic characteristics, respectively. Panel A (left column) of figure 4a and 4b plots the difference in coefficients between 1993 and 1983, panel B (middle column) plots the difference in coefficients between 2004 and 1993, and panel C (right column) plots the difference in coefficients between 2004 and 1983. We will investigate each variable's estimated effect on wages below.

4.1.1 Returns to Education

Figure 3a presents the returns to different levels of education. The intercept term represents the log wage distribution of the base group - primary educated workers belonging to non-Muslim, non-SC/ST group and employed in the manufacturing of machinery and transport equipment industry residing in the state of Tamil Nadu. The wages of the base group workers increase with the quantile across the distribution in all three years. As expected, wages increase with the level of education; in particular, a secondary or tertiary education increases the wage by a significant amount. While returns to different education levels are uniform across the distribution compared to the base group in 1983 and 1993; the returns to tertiary and secondary education are larger at higher quantiles in 2004.[17] So while in 1983

[17] Note that the approximately uniform returns are with respect to base group workers whose wage distribution is heterogeneous. The least squares miss the heterogeneity that comes through the wage distribution of base group as well as heterogeneity in returns to individual education levels compared with the base group.

and 1993, secondary and tertiary education contribute more to between group inequality; their contribution to within group inequality strengthened in 2004 (as returns become more heterogeneous).

Figure 4a presents the changes in returns to education over time. The intercept term shifted up in 1993 compared to 1983 (Panel A of Figure 4a). The base group workers are paid at an average approximately 30 percent more in real terms in 1993 compared to 1983. There is not much change in returns to different levels of education between 1983 and 1993 compared to the base group. Hence, all workers seem to be paid higher real wages in 1993 than in 1983. The intercept term shifted up again in 2004 compared to 1993 (Panel B of Figure 4a). The base group workers are paid approximately 20 percent more on average in real terms in 2004 compared to what they were paid in 1993. The returns to tertiary education increased nearly by 18 percent across the entire distribution and returns to secondary education also increased at higher quantiles. Consequently, the inequality increasing effect of secondary and tertiary education strengthened between 1993 and 2004.

Kijima (2006) also finds that the returns to primary and secondary education do not change much over time, while returns to tertiary education are stable up to 1993 and increase in 1999.[18] However, our findings suggest that return to secondary education also increased at the higher quantiles between 1993 and 2004.

4.1.2 Demographic Characteristics

Figure 3b presents the effects of demographic variables on wages in all three years. Female workers are paid significantly less across the entire wage distribution in all three years, though the disadvantage is less at higher quantiles. In addition to gender, there are few other demographic characteristics which play an important role in wage determination. It is generally argued that 'the caste system confines those from lower castes to a limited number of poorly paid, often socially stigmatized occupational niches from which there is little escape

[18] In Kijima (2006), primary level combines both primary and middle education levels.

. . . ' (Kabeer, 2002, p.3). Ethnicity is also often a source of exclusion - in India this translates into exclusion on the basis of religion and is largely applicable to the Indian Muslims (Das, 2003). Our findings also suggest disadvantage for lower castes and Muslims. The SC workers are paid less over the entire wage distribution in all three years.[19] However, the ST workers do not experience significant disadvantage in all three years. Muslim workers, who did not experience any disadvantage in 1983, are paid less in 1993 and 2004. Dutta (2006) also finds that belonging to the SC/ST or a Muslim group significantly decreases the wage received by regular workers in 1983, 1993, and 2004.

Figure 4b presents the changes in the effects of different demographic variables over time. The disadvantage faced by female workers did not decrease between 1983 and 1993. Rather it increased for female workers at lower quantiles. Between 1993 and 2004, the disadvantage faced by a female worker at lower quantiles decreased. However, at middle and higher quantiles, it increased marginally. This goes against the perception that increased competitiveness reduces female workforce disadvantage. There is not much change in the disadvantage faced by the SC workers between 1983 and 1993, but this disadvantage increased between 1993 and 2004. Interestingly, the SCs enjoy positive discrimination in public sector jobs, and one of the possible reasons for increase in disadvantage may be a decline in the share of public sector employment.

4.1.3 Industry Effects

Most of the industry dummies are statistically significant in all three years. The point estimate of industry effects are presented in Table 4 (for 1983) and Table 5 (for 2004). The coefficients are displayed as deviations from the employment-weighted average industry effect, and industries are ranked according to the magnitude of industry's effect at the mean

[19] Most of the previous studies combine the SCs and the STs together. However, while the exclusion and deprivation of the SCs is closely associated with institution of caste and untouchability, the STs isolation and exclusion however, is not related to caste or religion, but is based on their ethnic identity. Historically, the STs have been different from the mainland Indian society with a distinct culture, language, social organization, and economy practicing.

(given by OLS). The last row of Table 4 and 5 refers to joint significance of industry effects. In these tests, the null is always rejected, and we conclude that the industry matters for the determination of wages.

In 2004, a positive wage premium - compared to the average industry - is paid by sectors that either include capital intensive industries (e.g., mining, electricity, manufacturing of machinery, petroleum) or skill-intensive industries (e.g., computer, financial intermediation), while a negative wage premium is paid by sectors that include less capital intensive industries (e.g., light manufacturing such as the foodstuffs, tobacco, and textiles industries) or less skill-intensive industries (e.g., agriculture, hotel). These results are consistent with the findings of studies on other countries in which industries that are capital-intensive or skill-intensive (or both) have higher wage premia (Dickens and Katz 1987; Hasan and Chen 2003).

For most of the industries, there is no definite pattern in the industry wage premium across quantiles. The industries that pay a significant positive or negative wage premium tend to pay over the entire distribution. There have also been few changes in industrial structure, as reflected in the industry premium. Computer and business activity, which paid negative wage premium compared to the average industry in 1983, paid a significant positive wage premium in 2004. The textile industry paid a positive wage premium in 1983, but paid a significant negative wage premium in 2004. Similarly, post and telecommunication paid a significant wage premium over the average industry in 2004 versus no premium in 1983.

The marginal effects discussed heretofore vary across quantiles and over time. In addition, covariate distributions have changed over time (Table 1 and Table 3). To summarize the effects of changes in covariates and change in returns to the covariates on the overall change in wage distribution, we now turn to the MM (2005) decomposition.

4.2 Decomposition of Changes in Wage Distribution

Figure 5a, 5b, and 5c present the decomposition of wage differences between 1993 and 1983, between 2004 and 1993, and between 2004 and 1983, respectively, for quantile 5 to 95 with

95% confidence intervals. The confidence bounds are the 2.5% and 97.5% quantiles of the bootstrap distribution of the relevant statistic obtained by bootstrap with 1000 replications. Table 6 reports the results of the Oaxaca-Blinder and MM decompositions at select quantiles.

4.2.1 Wage Changes between 1983 and 1993

The Oaxaca-Blinder decomposition shows that real wages increased by 33 log points between 1983 and 1993, with the change in covariates explaining 9 log points and the change in coefficients explaining 24 log points. However, once we move beyond average, we find that real wages increased by 21 log points at the 10th percentile, 37 log points at the median, and 48 log points at the 90th percentile. Thus, increase in wages is heterogeneous across quantiles, with the increase being larger at the higher quantiles. Wage inequality increased while the real wages have been growing throughout the distribution, and the increase in wage inequality in this period is distributed over the entire wage distribution. In the presence of such heterogeneity in the increase in real wages, the Oaxaca-Blinder may hide important information.

The MM decomposition captures the heterogeneity in the covariate and the coefficient effects (Figure 5a): both are larger at higher quantiles. Change in both the covariates and the coefficients contribute to the actual evolution of real wages, and their effect is significantly different from zero (the confidence intervals do not include zero) at all of the estimated quantiles. However, the coefficient effect is quantitatively more important than the covariate effect at each of the estimated quantiles. Overall, the models work fairly well, as the residuals account for a relatively small part of the total change. The increase in the 90-10 gap in log wages is mostly due to change in observed prices (explain 69% of total change). The change in covariates explains 30% of the increase in the 90-10 gap.

4.2.2 Wage Changes between 1993 and 2004

The Oaxaca-Blinder decomposition shows that real wages increased by 20 log points between 1993 and 2004, and the change in the coefficients explains 25 log points, while the change in the covariates explains -6 log points. However, once we move beyond mean, we find that the growth of real wages between 1993 and 2004 has not only been heterogeneous but also has very different trend compared with growth of real wages between 1983 and 1993. Unlike the period 1983-1993, workers in most of the lower half of the distribution did not experience any increase in real wages. Real wages increased only for workers in the upper half of the wage distribution, and the increase is larger at higher quantiles. Thus wage inequality in this period increased with stable real wages at lower deciles of the wage distribution and increasing real wages at upper deciles. As a result, the increase in wage inequality between 1993 and 2004 is mostly concentrated in the upper half of the wage distribution.

Figure 5b presents the MM decomposition results. Unlike the 1980s, the change in covariates has a negative impact during 1993-2004 in most of the wage distribution. The coefficient effect is the main reason for the increase in wages, and it is positive throughout the distribution. The coefficient effect is larger at the higher quantiles. The positive coefficient effect just compensates the negative covariate effect until the median; and as a result, real wages are stable. After the median, the positive coefficient effect dominates the covariate effect. Consequently, workers in upper half of the wage distribution experience increase in real wages. The increase in prices explain majority of increase in the 90-10 gap in log wage (75% of the change), while change in observed covariates explains 27% of the increase (residuals account for -2%).

4.2.3 Wage Changes between 1983 and 2004

Overall, real wages increased by 52 log points between 1983 and 2004, and the Oaxaca-Blinder decomposition attributes 46 log points to the coefficient effect and 6 log points to the covariate effect. However, the increase in real wages is 27 log points at the 10^{th} percentile, 44 log points at the median and 91 log points at the 90^{th} percentile. Workers in the lower half of the wage distribution until the 40^{th} percentile experienced a uniform increase in real wages, while the increase in real wages was larger at higher quantiles after the 40^{th} percentile. The MM decomposition shows that the positive coefficient effect is the driving force behind the increase in real wages during this period, while the contribution of the covariate effect is either negligible or small. Most importantly, like the increase in real wages, the coefficient and the covariate effects are heterogeneous over the entire wage distribution. The increase in prices explains majority of the increase in the 90-10 gap in the log wage (70% of the change), while change in observed covariates explains 30% of the increase.

Our findings pertaining to the increase in wage differentials in 1990s is consistent with Kijima (2006) who also finds that increases in observed skill prices account for a dominant part of the increase in wage differentials (e.g., the 90^{th}-50^{th} gap) between 1993 and 1999. However, our results for the 1980s do not accord with the findings of Kijima (2006). Kijima (2006) finds that changes in the observed covariates played a dominant role in the increase in wage differentials between 1983 and 1987, and between 1987 and 1993. The differences in findings may arise for two reasons. First, the JMP decomposition, used by Kijima (2006), uses OLS coefficients in the decomposition which does not allow for heterogeneity in the coefficients across the distribution. Second, the sample used in Kijima (2006) differs from the sample used in this paper.[20]

To see why differences in the results arise, we also perform the JMP decomposition on our sample. The results are presented in Table 7. Between 1983 and 1993, the increase in returns dominates the changes in the covariates at the select quantiles reported. This is very much in conformity with our Oaxaca-Blinder and MM decompositions. However, differences in the results arise once we compute the reasons for the increase in wage differentials (e.g., the 90^{th}-50^{th}, the 50^{th}-10^{th}, and the 90^{th}-50^{th} gap as reported by Kijima (2006)). Since the JMP

[20] While our sample is restricted to full-time regular workers, Kijima (2006) uses full-time regular, casual, and self-employed workers. As wages are not reported for self employed workers, Kijima (2006) uses predicted wages for self-employed workers using OLS coefficients from regular salaried and casual workers (for whom wages are reported).

decomposition does not capture the heterogeneity in the coefficient effect, it understates the role of the coefficient effect in the increase in wage differentials. The MM decomposition captures this heterogeneity through the use of quantile regression. Hence, we believe that the coefficient effect played an important role in the increase in wage differentials in the 1980s also.

4.3 Decomposition of wage changes by gender

Wage structure of female workers in India has drawn little attention because of the low share of females in workforce. Both Dutta (2005) and Kijima (2006) concentrate on the wages of only male workers. Reilly and Dutta (2006) explore the gender pay gap in India in 1983, 1993, and 1999, and find that the gender pay gap exhibited a degree of stability over the period. However, they also concentrate on mean only. There is a growing literature looking at wage differentials between subgroups that go beyond simple mean comparisons. For example, several recent papers such as Albrecht, Bj¨orklund and Vroman (2003), Millimet and Wang (2006) look at gender gap across the entire wage distribution.

To examine how wages have evolved for the two genders, we decompose the wage changes over time for two genders separately. Figure 6 presents the decomposition results. Excluding female wage workers from the sample does not influence the main results (results obtained using both male and female samples) because of the low share of females in regular salaried workers (Panel A of Figure 6). The only difference from the main results is that male workers between the 30^{th} and 40^{th} percentiles did not see any increase in real wages between 1993 and 2004. In contrast, in the combined sample, workers between the 20^{th} and 50^{th} quantiles experienced no increase in real wages between 1993 and 2004.

However, the results are quite different for female workers when compared to our main results (male and female combined), or the results for male workers. Although the increase in female wages is very similar to the increase in male wages between 1983 and 1993, the driving force behind the increase in female wages differ from the driving force behind the increase in male wages (Panel B of Figure 6). In contrast to male workers (where the coefficient effect dominates the covariate effect across the distribution), the covariate effect dominates the coefficient effect until the 60^{th} percentile, and the coefficient effect dominates after the 60^{th} percentile. Moreover, the changes in wages for female workers between 1993 and 2004 indicate a very different trend than the changes in male wages. While female workers at lower and higher quantiles experienced an increase in real wages, female workers between the 20^{th} and 60^{th} percentiles actually experienced a decline in real wages. Focusing on average wages may mislead us to believe that female workers experienced increases in real wages in both decades; the average wages of female workers increased by 40 log points between 1983 and 1993, while it increased by 14 log points between 1993 and 2004. However, studying the entire distribution reveals that female workers in middle quantiles actually experienced a decline in real wages in the 1990s.

Our fitted model does not do a fair job in explaining the decrease in real wages, between 1993 and 2004, in the middle part of the wage distribution (as the residual is large). However, the decomposition suggests that the positive coefficient effect is driving the female wages and contribution of the covariates has been negative.

4.4 Decomposition of wage changes by education groups

The changes in the Indian economy over the last two decades may have had different effects on workers with different education levels. Hence, we decompose the wage changes experienced by different education groups to examine how different education groups have fared. The regressors included are age, age squared, and dummies for states, female, married, STs, SCs, Muslims and industries.

Figure 7 and Figure 8 plot the coefficient effect (increase in prices) for wage changes between 1983 and 1993, and between 1993 and 2004, respectively, for different education groups. Between 1983 and 1993, workers with all education levels experienced increases in prices paid (positive coefficient effect) except at extreme lower quantiles. Also, the coefficient effect experienced by different education group workers is approximately similar. However, between 1993 and 2004, the increase in prices paid to tertiary and secondary educated workers is much higher than the increase in prices paid to other education groups. Also, the increase in prices is heterogeneous across quantiles for tertiary and secondary educated workers. The coefficient effect is very small in most of the distribution for workers belonging to the below primary/illiterate group. This implies that skill premiums have been increasing in urban India during the last decade. Our finding of increasing skill premiums are consistent with Chamarbagwala (2006) and Kijima (2006).

5 Conclusion

The Indian economy has grown at a much faster rate in the past two decades compared to the 1950-80 period. The higher growth rate during the past two decades has also been associated with an increase in wage inequality. While wage inequality in India is well studied, all such studies (to our knowledge) focus on averages, relying mostly on OLS estimates. We find that such a narrow focus does not depict the complete picture in a heterogeneous society like India. We find that real wages increased across the entire wage distribution between 1983 and 1993 and the increase is larger at the higher quantiles. However, real wage remained stable in most of the lower half of the distribution between 1993 and 2004 while it increased in the upper half of the wage distribution with the increase being larger at the higher quantiles. The findings do not change when we restrict our sample to male workers only. While female workers experienced similar changes in real wages between 1983 and 1993, between 1993 and 2004 their wages declined in real terms in the middle of the wage distribution while increasing at the lower and higher quantiles.

Further decomposing the changes in wage, we find that the increase in real wage over the last two decades is driven mostly by the increases in prices paid. Between 1983 and 1993, both the coefficient and the covariate effects contributed to increase in wages, though the coefficient effect was more important than the covariate effect. However, between 1993 and 2004, the contribution of the covariate effect was either negative or negligible while almost all of the increase in wages has been due to the positive coefficient effect.

Our findings also suggest that returns to many characteristics (education levels, demographic characteristics) are heterogeneous and the least squares method fails to capture this heterogeneity. The return to tertiary education not only increased by almost 20 percent between 1993 and 2004 but also became more heterogeneous in 2004. This reinforced the inequality increasing effects of tertiary education. Moreover, between 1983 and 1993, workers with all levels of education experienced increases in prices paid; between 1993 and 2004, the increase in prices paid to tertiary and secondary educated workers is much higher than the increase for workers with other levels of education. Also, this increase in prices for tertiary and secondary educated workers is larger at the higher quantiles.

The findings of the paper suggest that wage inequality in urban India is not likely to decrease and in fact will probably increase further in the near future as the share of tertiary educated workers in the workforce increases. However, in the long run, increases in the skill premium are expected to stimulate further increases in human capital investment (Topel, 1999), and increases in the number of college graduates may decrease the wage inequality in the long run as South Korea experienced in the 1970s and 1980s (Kim and Topel, 1995).

This paper raises some new questions which have policy and welfare implications. Why did the real wages of female workers decline in the middle of the wage distribution between 1993 and 2004? Why did the structural reforms during the 1990s affect male and female wages differently? The female wage structure in India has not drawn much interest because of their low share in workforce. However, given the significant differences in behavior of wage changes for two genders during 1990s, additional research regarding the deeper sources underlying the changes in gender equality should prove insightful. Also, the issue of gender gap should be revisited as the average gender gap explained by existing literature may not be satisfactory given the high degree of heterogeneity in wages paid and changes in wages over time.

References

[1] Ahluwalia, M S. (2002), Economic Reforms in India since 1991: Has Gradualism Worked?, *Journal of Economic Perspectives*, 16(3), 67-88.

[2] Albrecht, J., Bjorklund, A. and Vroman, S. (2003), Is there a Glass Ceiling in Sweden?, *Journal of Labor Economics*, 21, 145-177.

[3] Albrecht, J., Vuuren, A. and Vroman, S. (2007), Counterfactual Distributions with Sample Selection Adjustments: Econometric Theory and an Application to the Netherlands, *Georgetown University Working Paper.*

[4] Arbache, J.S., Dickerson, A. and Green, F. (2004), Trade Liberalisation and Wages in Developing Countries, *The Economic Journal*, 114, F73–F96.

[5] Autor, D.H., Katz, L.F. and Kearney, M.S. (2005), Rising Wage Inequality: The Role of Composition and Prices, *NBER Working Paper*, 11628.

[6] Banerjee, A. and Piketty, T. (2003), Indian Top Incomes, 1956-2000, *CEPR Discussion Paper*, 4137.

[7] Buchinsky, M. (1994), Changes in the U.S. Wage Structure 1963-1987: An Application of Quantile Regression, *Econometrica*, 62(2), 405-58.

[8] Buchinsky, M. (1998), The Dynamics of changes in the Female Wage Distribution in the USA: A Quantile Regression Approach, *Journal of Applied Econometrics*, 13, 1-30.

[9] Chamarbagwala, R. (2006), Economic Liberalization and Wage Inequality in India, *World Development*, 34(12), 1997-2015.

[10] Das, M. B. (2003), Ethnicity and Social Exclusion in Job Outcomes in India: Summary of Research Findings, *unpublished paper*, World Bank Institute.

[11] Dickens, W T. and Katz, L. F. (1987), Inter-Industry Wage Differences and Theories of Wage Determination, *NBER Working Paper*, 2271.

[12] DiNardo, J., Fortin, N. and Lemieux, T. (1996), Labor Market Institutions and the Distribution of Wages, 1973–1992: a Semiparametric Approach, *Econometrica*, 64, 1001–1044.

[13] Dutta, P.V. (2005), Accounting for Wage Inequality in India, *Poverty Research Unit at Sussex Working Paper*, 29.

[14] Dutta, P.V. (2006), Returns to Education: New Evidence for India, 1983-1999, *Education Economics*, 14, 431-451.

[15] Fitzenberger, B. and Kurz, C. (2003), New Insights on Earning trends across Skill groups and Industries in West Germany, *Empirical Economics*, 28, 479-514.

[16] González, X. and Miles, D. (2001), Wage Inequality in a Developing Country: Decrease in Minimum Wage or Increase in Education Returns, *Empirical Economics*, 26, 135-148.

[17] Government of India. (1992), Economic Survey, 1991-92.

[18] Government of India. (2006), Social, Economic and Education Status of the Muslim community of India, Report submitted by the Prime Minister's High Level Committee, New Delhi.

[19] Hasan, R. and Chen, L. (2003), Trade and Workers: Evidence from Philippines, *Economic Study Area Working Paper*, 61.

[20] Joshi, V. and Little, I.M.D. (1994), India: Macroeconomics and Political Economy, 1964–1991, *World Bank*, Washington, D.C.

[21] Juhn, C., Murphy, K. and Pierce, B. (1993), Wage Inequality and the Rise in Returns to Skill, *Journal of Political Economy*, 101, 410-442.

[22] Kabeer, N. (2002), Safety Nets and Opportunity Ladders: Addressing Vulnerability and Enhancing Productivity in South Asia, *Overseas Development Institute Working Paper,* 159.

[23] Katz, L. and Autor, D. (1999), Changes in the Wage structure and Earnings inequality, in: Ashenfelter,O., Card, D. (Eds.), Handbook of Labor Economics, Elsevier, Amsterdam, 1463-1555.

[24] Kim, D. and Topel, R. (1995), Labor Markets and Economic Growth: Lessons from Korea's Industrialization, 1970–1990. In: Freeman, R., Katz, L. (Eds.), Differences and Changes in Wage Structures. University of Chicago Press for NBER, Chicago, pp. 227–264.

[25] Kijima, Y. (2006), Why did Wage Inequality Increase? Evidence from Urban India 1983-99, *Journal of Development Economics*, 81(1), 97-117.

[26] Koenker, R. and Bassett, G. (1978), Regression Quantiles, *Econometrica*, 46, 33-50.

[27] Leibbrandt, M., Levinsohn, J. and McCrary, J. (2005), Incomes in South Africa Since the Fall of Apartheid, *NBER Working Paper*, 11384.

[28] Lemieux, T. (2006), Post Secondary Education and Increasing Wage Inequality, *American Economic Review*, 96(2), 1-23.

[29] Lemieux, T. (2007), Changing Nature of Wage Inequality, *University of British Columbia*, Mimeo.

[30] Machado, J. and Mata. J. (2001), Earning functions in Portugal 1982-1994: Evidence from Quantile Regressions, *Empirical Economics*, 26, 115-34.

[31] Machado, J. and Mata. J. (2005), Counterfactual Decomposition of Changes in Wage Distributions using Quantile Regression, *Journal of Applied Econometrics*, 20, 445-65.

[32] Melly, B. (2005), Decomposition of Differences in Distribution using Quantile Regression, *Labour Economics*, 12, 577-590.

[33] Millimet, D. and Wang, Le. (2006), A Distributional Analysis of the Gender Earnings Gap in Urban China, *Contributions to Economic Analysis and Policy*, 5(1), Article 5.

[34] Oaxaca, R. (1973), Male-Female Differentials in Urban Labor Markets, *International Economic Review*, 14, 693-709.

[35] Oaxaca, R. and Ransom, M. (1994), On Discrimination and the Decomposition of Wage Differentials, *Journal of Econometrics*, 61, 5-21.

[36] Panagariya, A. (2004), India in 1980s and 1990s: A Triumph of Reforms, *IMF Working Paper*, 04/43.

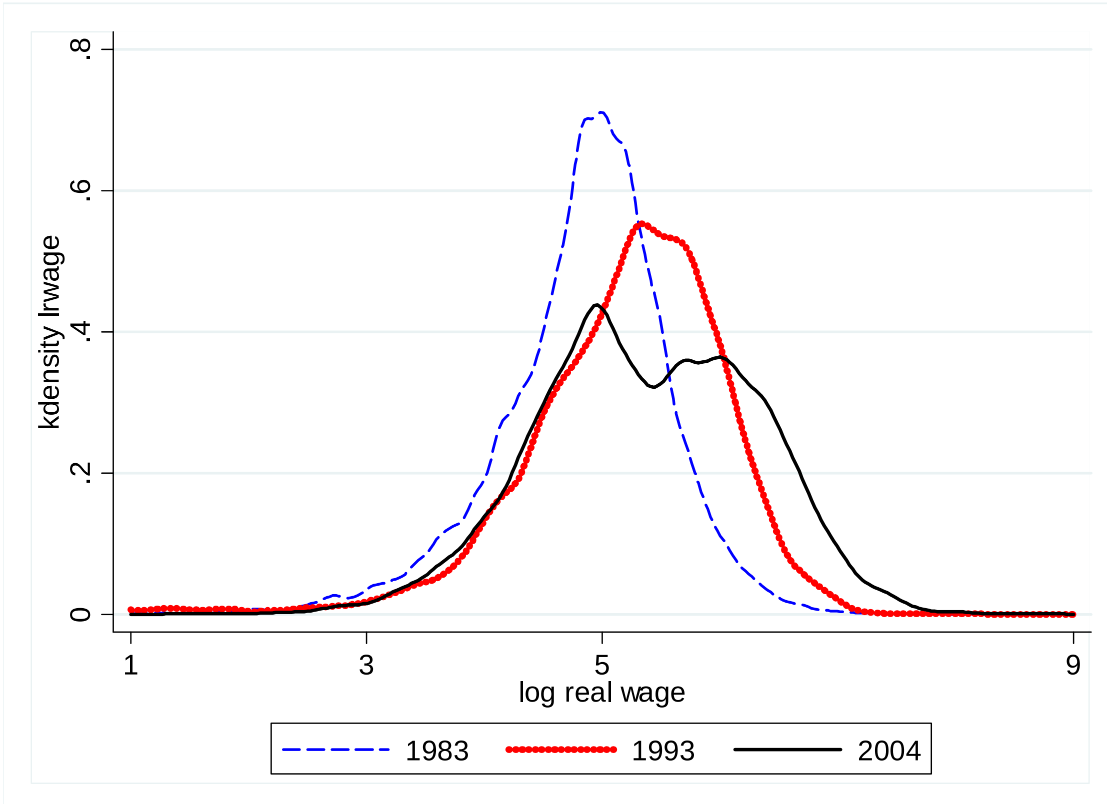
[37] Reilly, B. and Dutta, P.V. (2005), The Gender Pay Gap and Trade Liberalization: Evidence for India, *Poverty Research Unit at Sussex Working Paper*, 32.

[38] Topel, R. (1999), Labor Markets and Economic Growth. In: Ashenfelter, O., Card, D. (Eds.), Handbook of Labor Economics. Elsevier, Amsterdam, pp. 2943– 2984.

[39] Wood, A. (1997), Openness and Wage Inequality in Developing Countries: The Latin American Challenge to East Asian Conventional Wisdom, *World Bank Economic Review*, 11(1), 33-57.

[40] Zanchi, L. (1998), Inter-industry Wage Differentials in Dummy Variable Models, *Economic Letters*, 60, 297-301.

Figure 1: Kernel density of log real wage



Note: Gaussian kernel is used. The chosen width is the width that would minimize the mean integrated squared error.

Figure 2: Cumulative distribution of log real wage

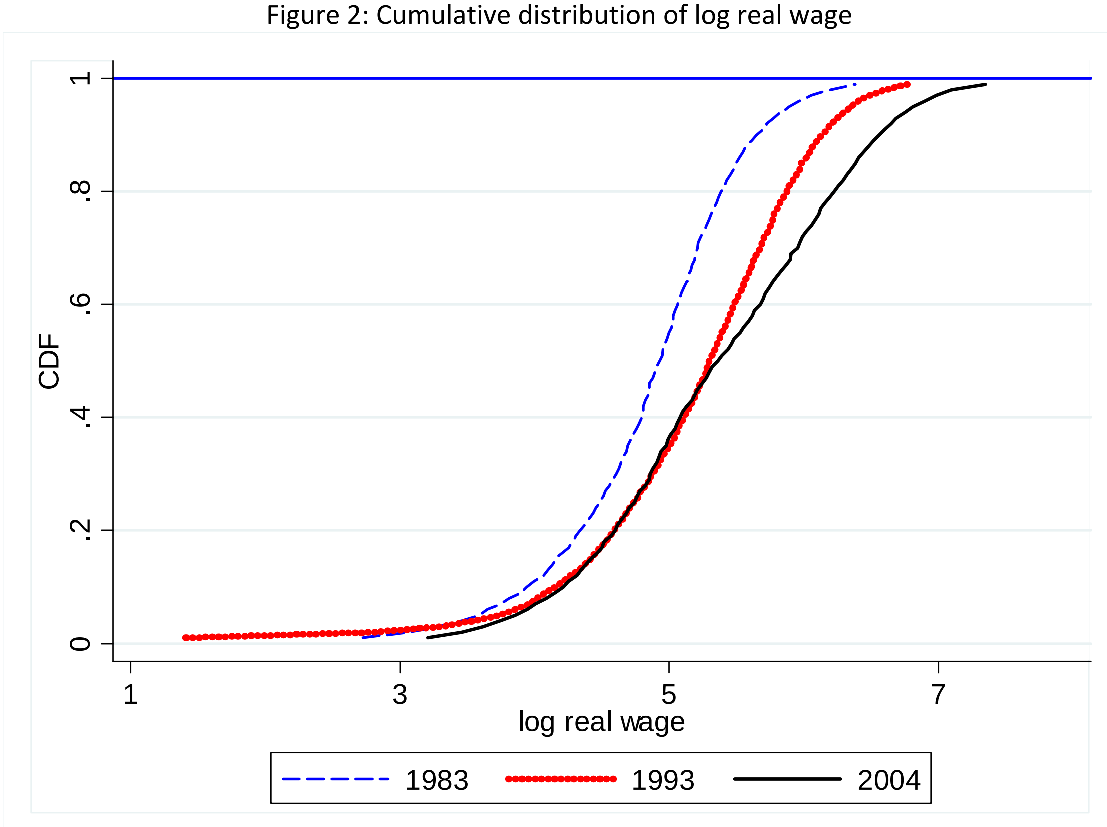


Figure 3a: Regression Coefficients for different stages of education

|--------------Panel A------------------| |--------------Panel B-------------- | |--------------Panel C----------------|

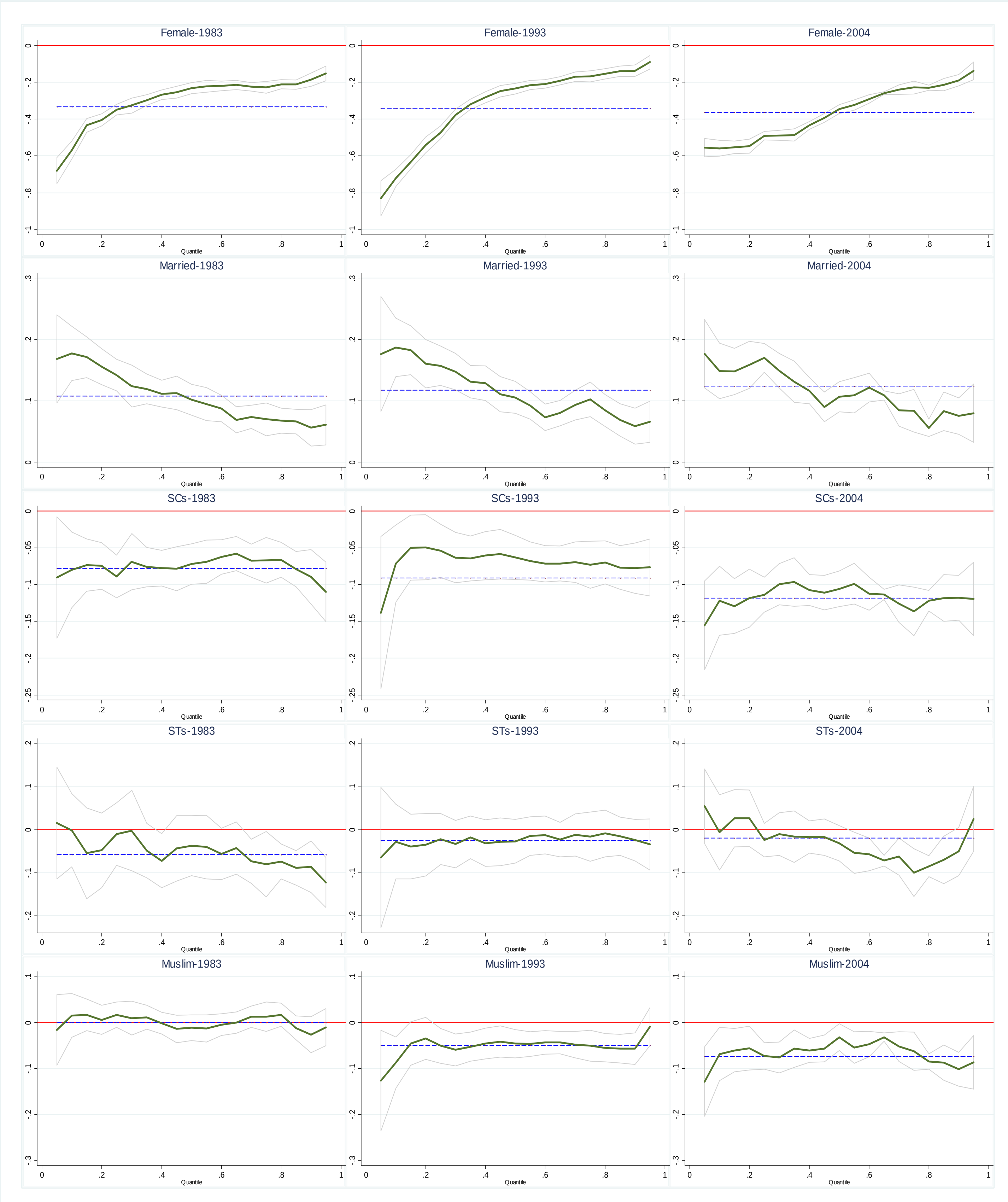
"1983" "1993" "2004"

Notes: (1) Below Primary/Illiterate is excluded education category. (2) Dashed horizontal line refers to OLS estimate. (3) Y-axis scale differs for different levels of education.

Figure 3b: Regression Coefficients for demographic variables

|--------------Panel A------------------| |--------------Panel B-------------- | |--------------Panel C----------------|

"1983" "1993" "2004"

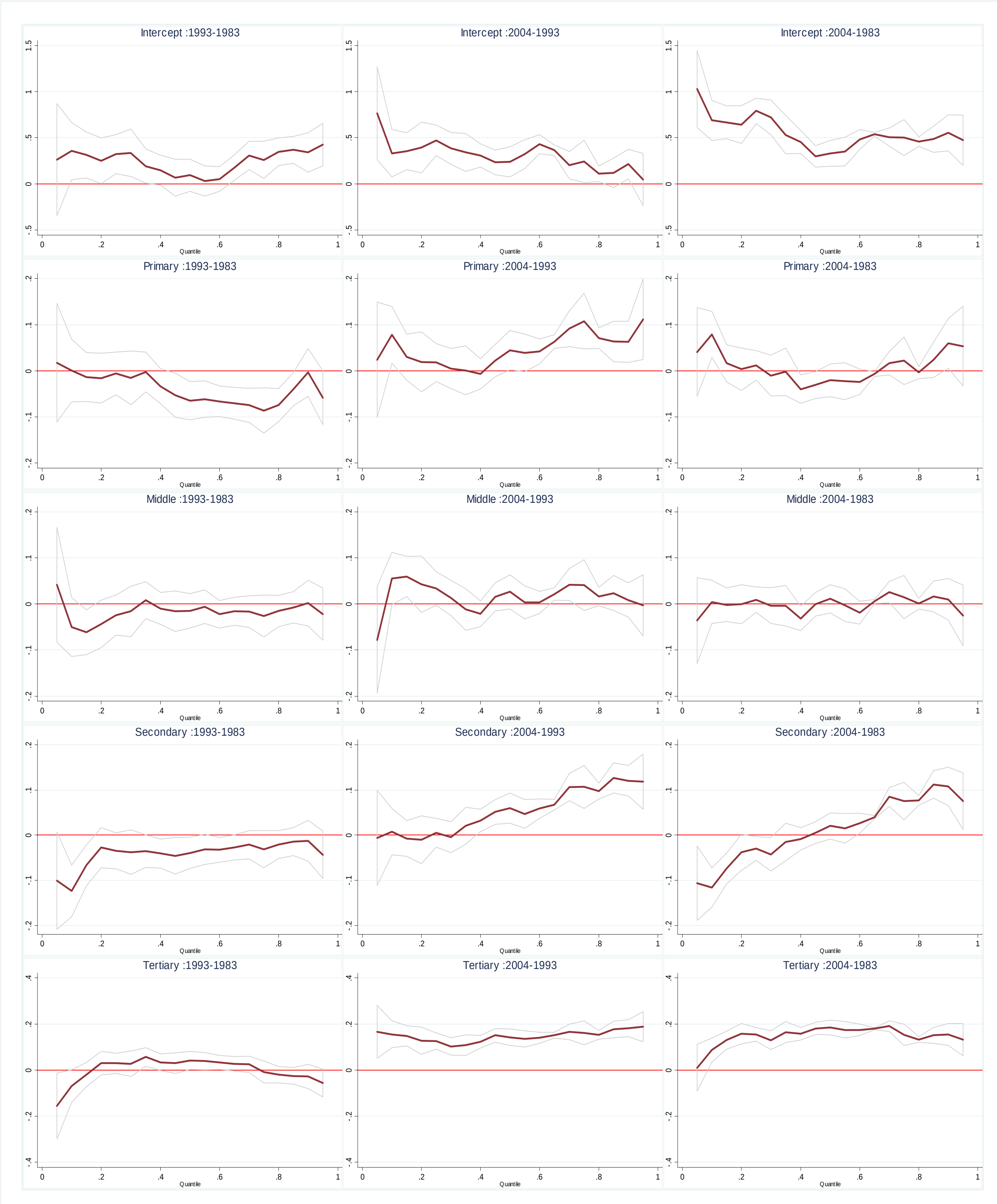


Notes: (1) Solid horizontal line refers to zero. (2) Dashed horizontal line refers to OLS estimate. (3) Y-axis scale differs for different characteristics.

Figure 4a: Changes in Regression Coefficients of different levels of education

|---------------Panel A-------------------| |---------------Panel B--------------- | |---------------Panel C----------------|

"1993-1983" "2004-1993" "2004-1983"

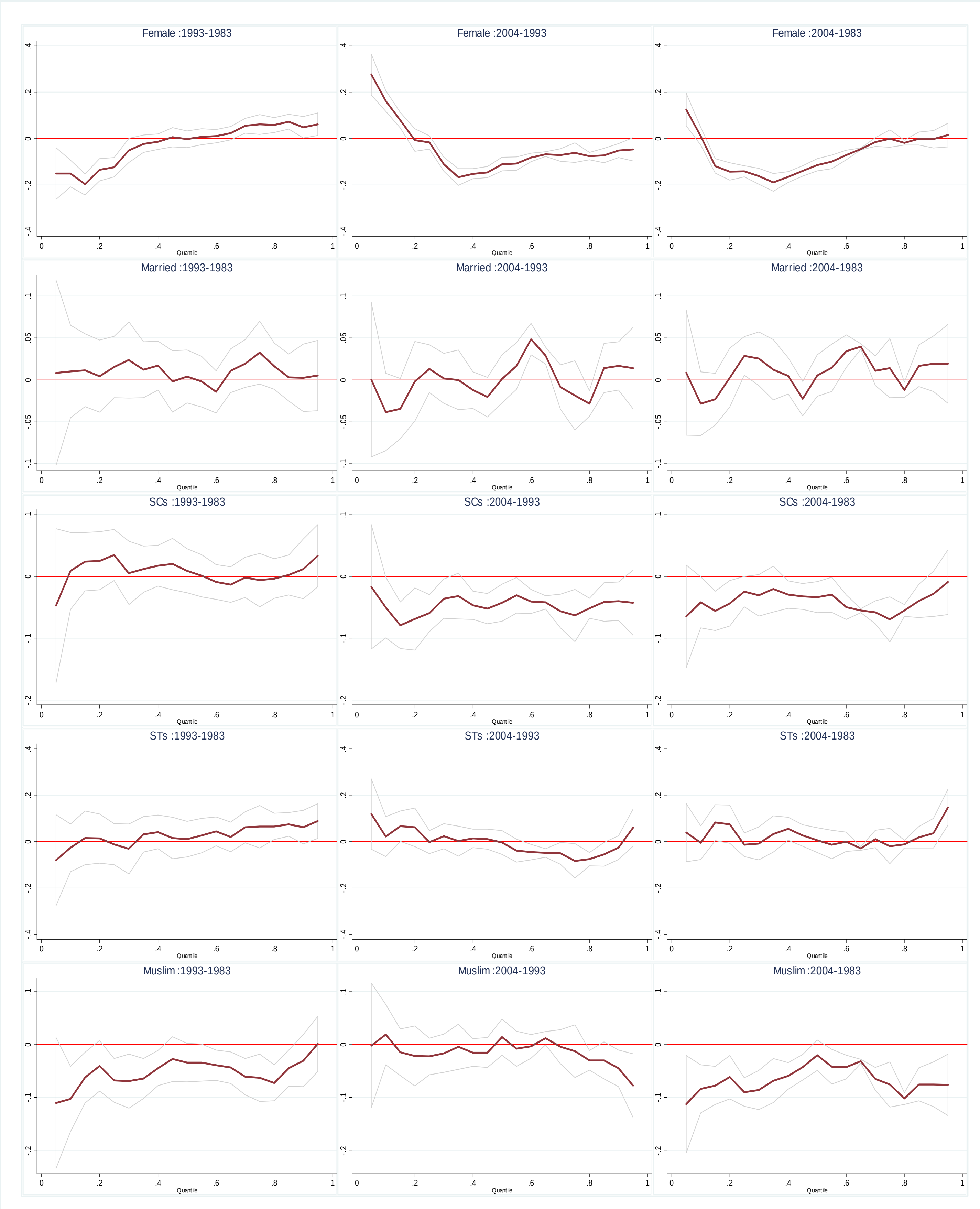


Note: Solid horizontal line refers to zero.

Figure 4b: Changes in Regression Coefficients of demographic variables.

|---------------Panel A-------------------| |---------------Panel B--------------- | |---------------Panel C-----------------|

"1993-1983" "2004-1993" "2004-1983"



Note: Solid horizontal line refers to zero.

Figure 5a: Decomposition of difference in wages between 1983 1993

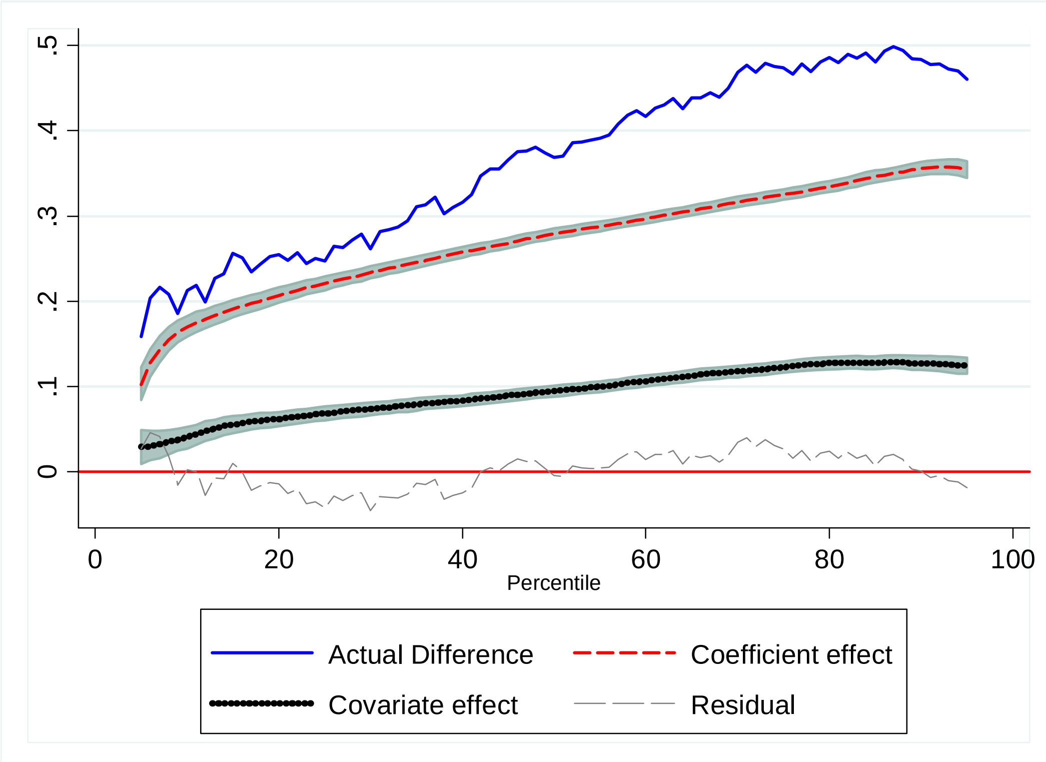
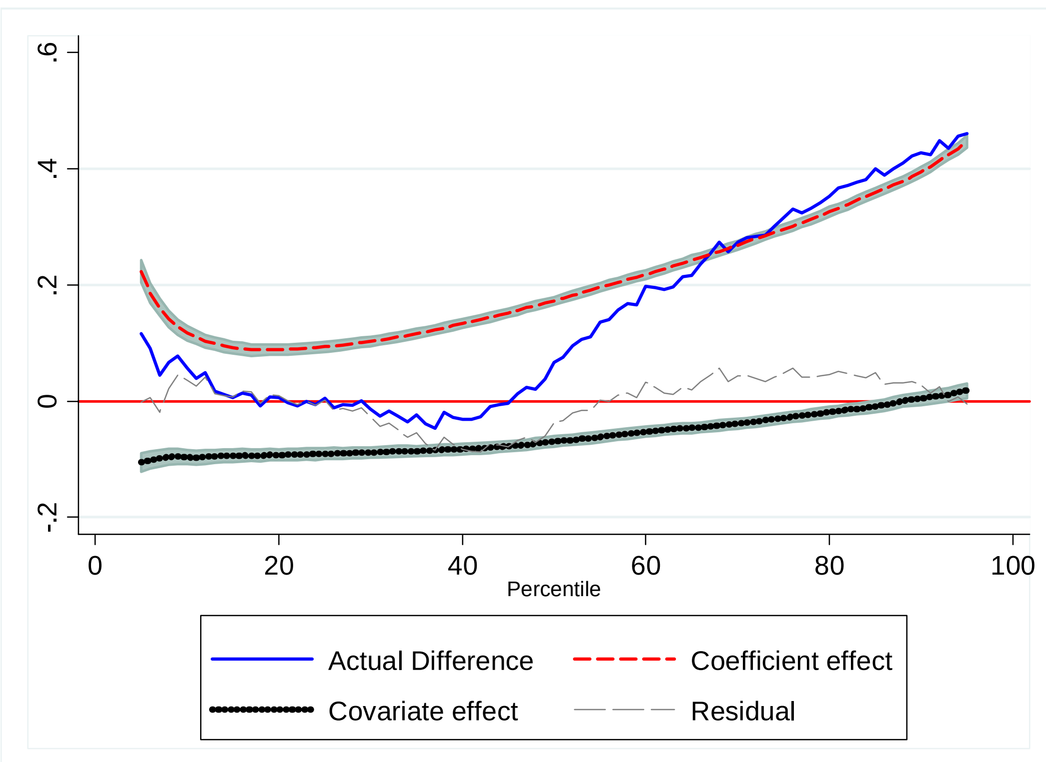


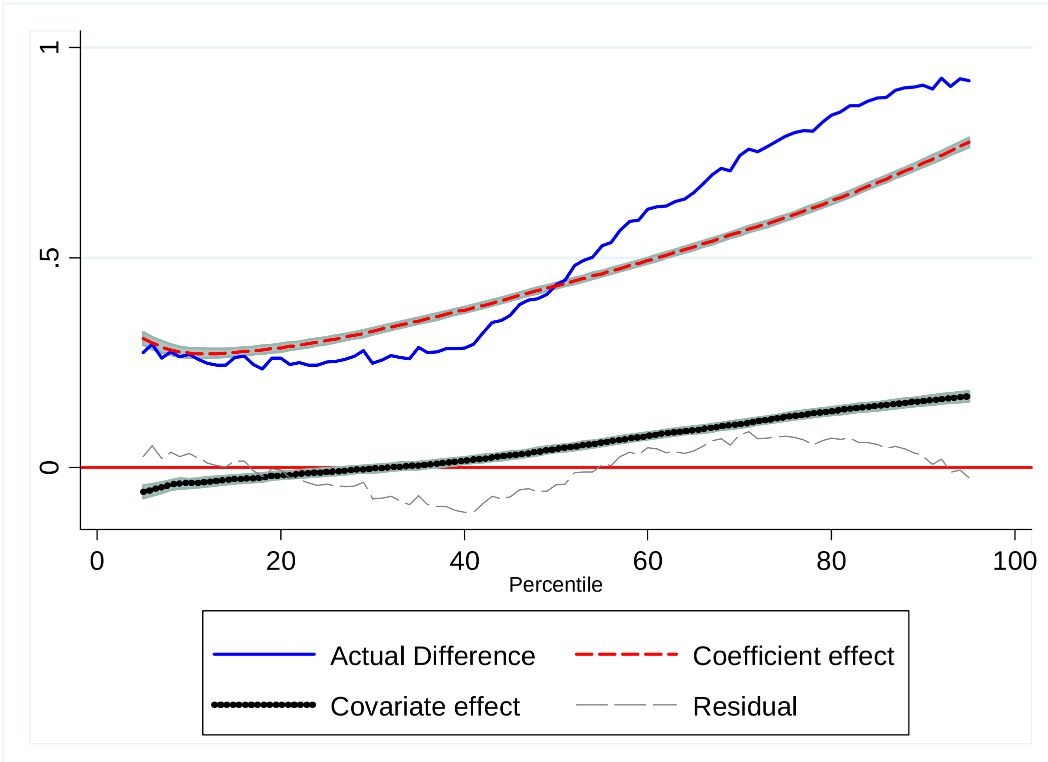
Figure 5b: Decomposition of difference in wages between 2004 and 1993



Continued……..

Figure 5 Continued……….

Figure 5c: Decomposition of difference in wages between 2004 and 1983



Notes: (1) Actual difference is the difference of actual empirical densities of year 1 from year 0 at each quantile, i.e., α {f(w(1))} – α {f(w(0))}. Year 1 refers the current year and year 0 refers previous year.

(2) Residual is the difference of difference in actual empirical densities and difference in fitted densities at each quantile, i.e, {α(f (w(1))) - α(f(w(0)))} – {α(f*(w(1))) - α(f*(w (0)))}.

(3) Coefficient effect is difference between the counterfactual density (with prices of year1 and covariates of year 0) and the fitted density for year 0, i.e., α{f *(w(1);x(0))} – α{f*(w(0))}.

(4) Covariate effect is difference of fitted density of year 1 and counterfactual density (with prices of year1 and covariates of year zero), i.e., α{f*(w(1))} – α{f*(w(1);x(0))}.

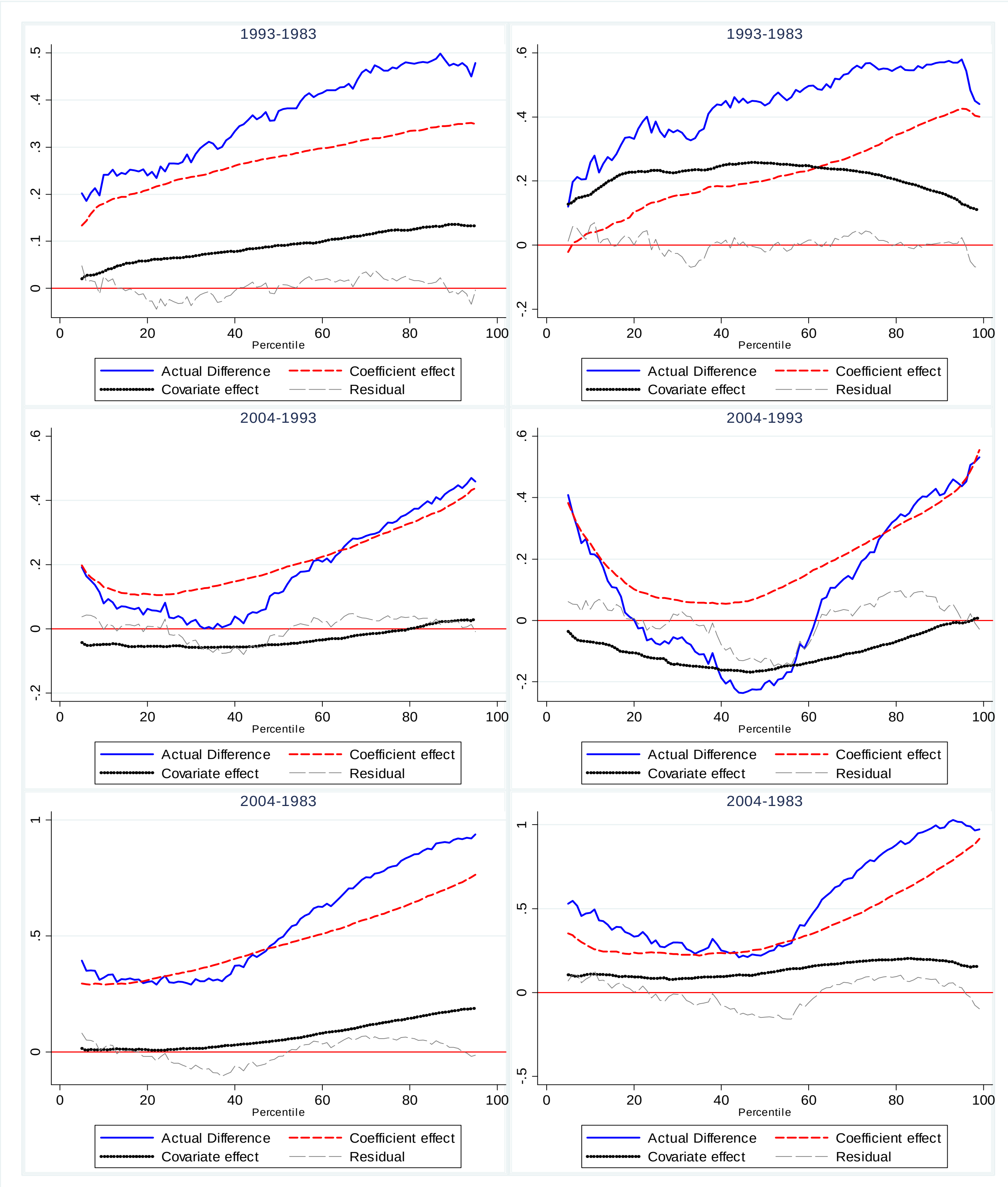
(5) The shaded area around coefficient and covariate effect refers to 95% confidence interval. The confidence bounds are the quantiles 2.5% and 97.5% of the distribution of the relevant statistic obtained by bootstrap with 1000 replications.

See text for further details.

Figure 6: Decompostition of Wage Change over time by Gender

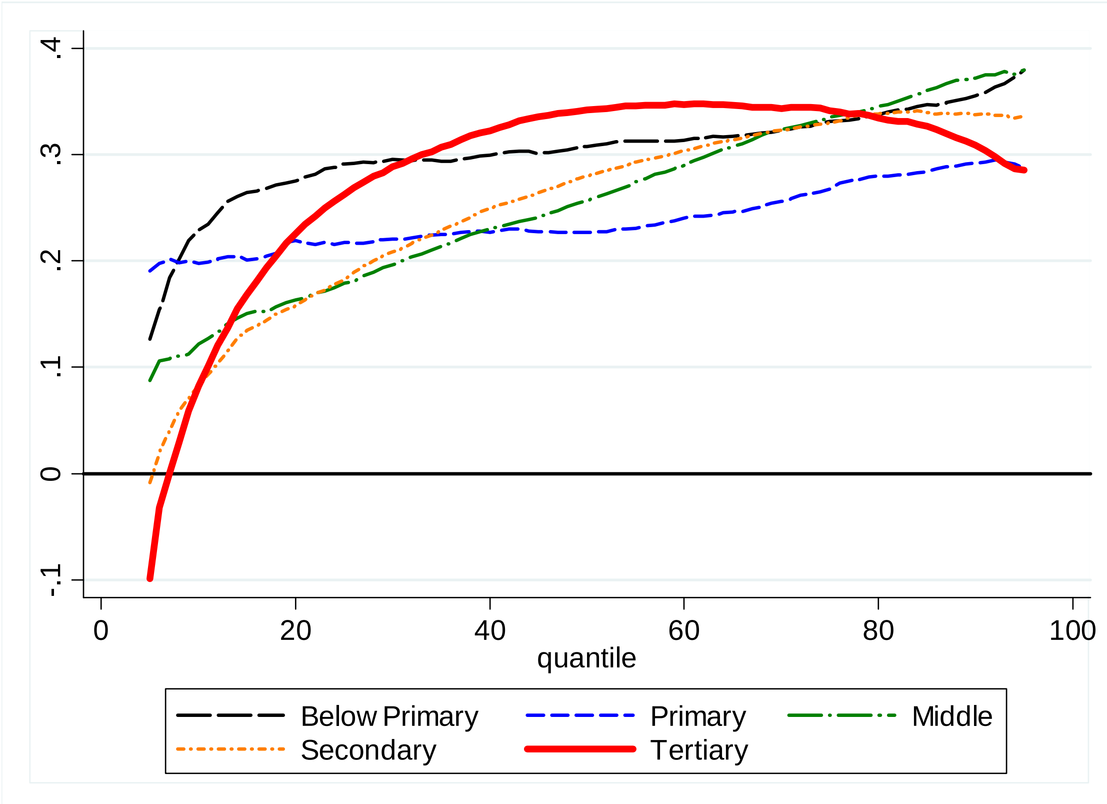
|-----------------------Panel A-----------------------------| |-----------------------------Panel B---------------------|

Male **Female**



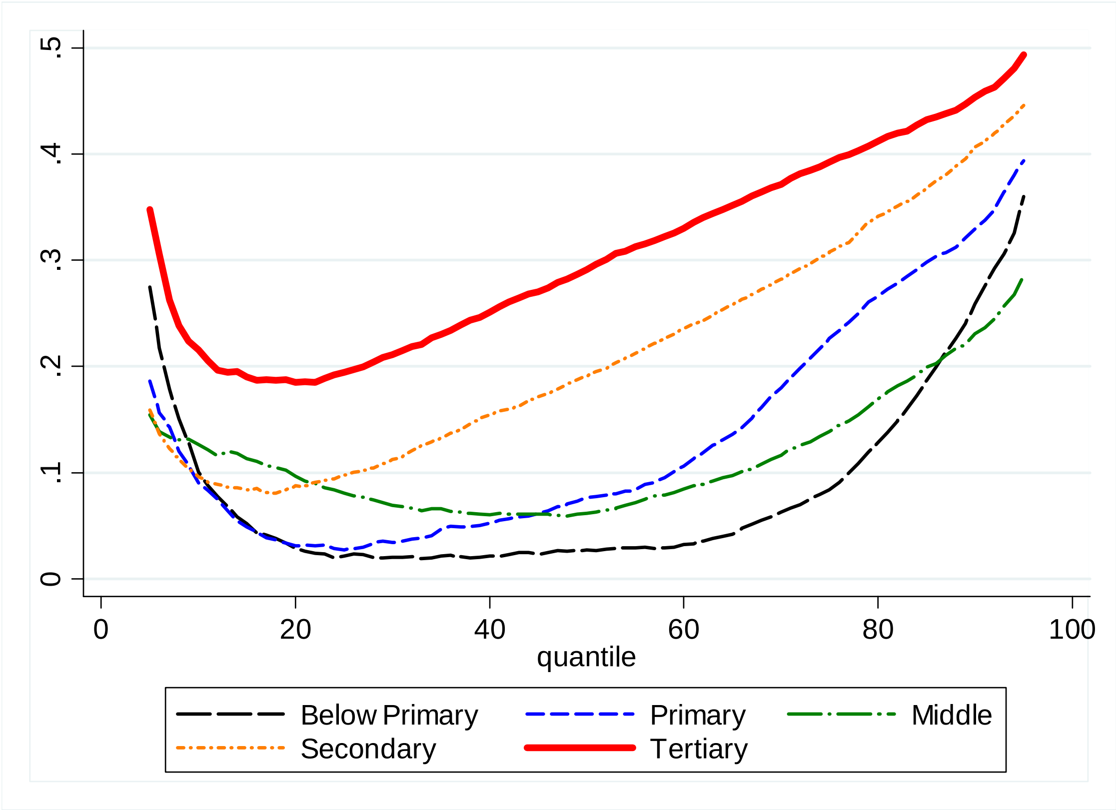
Note: See Notes below Figure 5.

Figure 7: Price Differential by educational attainment of workers between 1993-1983



Note: Price differential for each education category workers is the coefficient effect (decomposing the wage differences between two years for each education group separately) that is estimated by α{f*(w(1);x(0))}-α{f*(w(0))}. See text for further details.

Figure 8: Price Differential by educational attainment of workers between 2004-1993



Note: Price differential for each education category workers is the coefficient effect (decomposing the wage differences between two years for each education group separately) that is estimated by α{f*(w(1);x(0))}-α{f*(w(0))}. See text for further details.

Table 1: Descriptive statistics of sample

Variable\year	1983	1993-94	2004-05
Real wage	161.41	241.04	324.26
	(140.35)	(182.46)	(371.70)
Log real wage	4.85	5.18	5.38
	(0.71)	(0.92)	(0.92)
Percentile Differential in log real wage:			
90-10	1.71	1.98	2.35
90-50	0.72	0.84	1.20
75-50	0.37	0.47	0.72
75-25	0.81	1.03	1.34
50-25	0.44	0.56	0.62
50-10	0.98	1.14	1.15
Age	36.09	37.35	37.06
	(9.79)	(9.69)	(10.20)
Education levels (Proportion in each level)			
Illiterates	0.14	0.11	0.09
Below Primary	0.09	0.08	0.06
Primary	0.13	0.10	0.10
Middle	0.17	0.15	0.15
Secondary	0.28	0.31	0.32
Tertiary	0.18	0.26	0.29
Demographic variables (proportion in working population)			
Female	0.13	0.15	0.19
Married	0.81	0.83	0.77
Scheduled Castes	0.12	0.11	0.15
Scheduled Tribes	0.03	0.03	0.03
Muslims	0.09	0.08	0.09
Sample Size	24,059	24,907	23,276

Notes: 1) Standard deviation in parenthesis.

2) Real wages is in Indian Rupees at 1984-85 prices.

Table 2: Distribution of different education level workers across quintile

Year\Quintile	1	2	3	4	5
	Illiterate/Below Primary				
1983	36.19	28.69	18.18	12.38	4.56
1993	38.01	27.20	20.67	11.22	2.90
2004	47.98	23.34	16.95	9.55	2.18
	Primary				
1983	25.87	27.48	21.18	18.02	7.45
1993	30.68	27.75	25.71	12.29	3.57
2004	33.79	25.51	21.87	15.10	3.73
	Middle				
1983	19.11	24.39	22.65	21.55	12.3
1993	23.29	24.81	26.34	18.35	7.21
2004	25.32	25.63	23.17	18.88	7.00
	Secondary				
1983	9.09	13.79	19.14	27.26	30.71
1993	12.88	15.14	21.84	28.23	21.90
2004	14.14	15.37	19.73	27.03	23.73
	Tertiary				
1983	4.31	6.33	11.46	22.90	55.00
1993	7.87	6.62	11.14	23.51	50.86
2004	6.31	7.14	13.18	24.81	48.57

Table 3: Shares of different industries in regular employment in urban India

Industry Name	1983	1993-94	2004-05	change in share between 2004-05 and 1983
Agriculture, hunting, forestry and fishing	1.25	1.02	0.85	-0.40
Mining and quarrying	2.37	2.31	1.70	-0.67
Manufacture of food, beverage and tobacco products	2.87	2.33	1.98	-0.89
Manufacture of textiles, leather, fur, wearing apparel and footwear	8.94	7.42	8.43	-0.51
Manufacture of wood and wood products	0.70	0.46	2.32	1.62
Manufacture of paper, paper products, printing and publishing	1.38	1.34	1.61	0.23
Manufacture of chemicals, rubber, plastic, petroleum and coal products	2.79	3.58	3.01	0.22
Manufacture of nonmetallic mineral products	1.19	0.89	0.77	-0.42
Manufacture of basic metals, metal products and metal parts	3.67	3.58	2.91	-0.76
Manufacture of machinery, transport equipment and parts	5.55	6.38	4.47	-1.08
Electricity, gas, steam, water works and water supply	2.23	2.48	1.75	-0.48
Construction	1.61	1.76	1.34	-0.27
Wholesale and retail trade-repair of motor vehicles and personal household goods	6.31	7.52	10.22	3.91
Hotels and restaurants	1.62	1.54	2.42	0.80
Transport and storage	10.18	9.55	8.56	-1.62
Post and telecommunications	1.34	1.33	1.90	0.56
Financial intermediation	3.66	4.83	4.50	0.84
Real estate, renting	0.73	0.10	0.34	-0.39
Computer and related activities /professional business activity	0.20	1.54	3.53	3.33
Public administration and defense	22.89	21.64	14.49	-8.40
Education and R&D	9.09	9.68	11.51	2.42
Health, social work	3.49	2.87	3.78	0.29
Sanitation related activities	0.83	0.84	0.30	-0.53
Other social activity	5.07	5.00	7.30	2.23
	100	100	100	

Table 4: Industry Affects, 1983

Industry	OLS	Percentile 10th	25th	50th	75th	90th
Mining and quarrying	0.37	0.40	0.37	0.33	0.31	0.39
Financial intermediation	0.28	0.35	0.31	0.27	0.26	0.28
Electricity, gas, steam, water works, and water supply	0.20	0.24	0.23	0.20	0.17	0.12
Health, social work	0.13	0.15	0.15	0.14	0.07	0.03
Manufacture of chemicals, rubber, plastic, petroleum, and coal products	0.13	0.10	0.09	0.13	0.18	0.19
Manufacture of basic metals, metal products, and metal parts	0.11	0.03	0.05	0.09	0.14	0.18
Transport and storage	0.09	0.15	0.12	0.09	0.05	0.05
Manufacture of machinery and transport equipment and parts	0.08	0.01	0.05	0.12	0.14	0.15
Public administration and defense	0.07	0.14	0.11	0.07	0.02	0.00
Manufacture of textiles, leather, fur, wearing apparel, and footwear	0.06	0.03	0.07	0.07	0.08	0.05
Post and telecommunications	0.01	0.07	0.09	0.03	0.00	-0.12
Sanitation related activities	-0.01	0.09	0.10	0.04	-0.03	-0.08
Education and R&D	-0.03	-0.04	0.02	0.00	-0.01	-0.09
Manufacture of wood and wood products	-0.06	0.06	-0.07	-0.11	-0.09	-0.05
Construction	-0.06	-0.13	-0.18	-0.11	0.02	0.10
Manufacture of nonmetallic mineral products	-0.08	-0.18	0.04	-0.03	-0.11	-0.13
Manufacture of paper, paper products, printing, and publishing	-0.13	-0.09	-0.18	-0.15	-0.09	-0.07
Real estate, renting	-0.17	-0.36	-0.30	-0.08	0.01	-0.03
Hotels and restaurants	-0.20	-0.27	-0.23	-0.20	-0.23	-0.16
Computer and related activities /professional business activity	-0.23	-0.18	-0.16	-0.21	-0.19	-0.28
Manufacture of food, beverage, and tobacco products	-0.27	-0.27	-0.31	-0.30	-0.28	-0.17
Wholesale and retail trade-repair of Motor vehicles and personal household goods	-0.27	-0.31	-0.33	-0.30	-0.24	-0.19
Agriculture, hunting, forestry, and fishing	-0.36	-0.55	-0.45	-0.31	-0.29	-0.21
Other social activity	-0.58	-0.75	-0.78	-0.63	-0.44	-0.34
Joint Significance of industry effects (p- values)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)

Notes: (1) The coefficients are expressed as deviation from employment weighted average industry effect. (2) Industries are ranked according to the magnitude of industry effects at the mean (given by OLS).

Table 5: Industry Effects, 2004-05

	Percentile					
Industry	**OLS**	**10th**	**25th**	**50th**	**75th**	**90th**
Mining and quarrying	0.68	0.70	0.86	0.69	0.59	0.58
Electricity, gas, steam, water works, and water supply	0.43	0.58	0.54	0.43	0.36	0.36
Financial intermediation	0.34	0.31	0.38	0.38	0.37	0.24
Public administration and defense	0.28	0.45	0.38	0.31	0.23	0.16
Computer and related activities /professional business activity	0.24	0.03	0.03	0.16	0.44	0.47
Post and telecommunications	0.21	0.02	0.16	0.24	0.29	0.37
Transport and storage	0.12	0.04	0.11	0.14	0.13	0.11
Health, social work	0.11	-0.02	0.05	0.10	0.18	0.23
Manufacture of machinery and transport equipment and parts	0.09	0.11	0.06	0.00	0.12	0.18
Manufacture of chemicals, rubber, plastic, petroleum, and coal products	0.05	0.11	0.01	0.04	0.05	0.12
Education and R&D	0.03	-0.20	-0.01	0.17	0.14	0.08
Sanitation related activities	0.00	-0.29	0.08	0.10	0.08	0.14
Manufacture of basic metals, metal products, and metal parts	-0.04	-0.08	-0.09	-0.13	-0.03	0.17
Manufacture of wood and wood products	-0.08	0.00	-0.01	-0.20	-0.17	-0.12
Construction	-0.08	-0.19	-0.21	-0.11	0.10	0.04
Manufacture of nonmetallic mineral products	-0.13	-0.07	-0.02	-0.30	-0.28	0.01
Hotels and restaurants	-0.16	-0.10	-0.18	-0.18	-0.22	-0.19
Manufacture of textiles, leather, fur, wearing apparel, and footwear	-0.16	-0.02	-0.12	-0.18	-0.26	-0.31
Manufacture of paper, paper products, printing, and publishing	-0.17	-0.15	-0.17	-0.16	-0.22	-0.09
Agriculture, hunting, forestry, and fishing	-0.26	-0.31	-0.27	-0.27	-0.30	-0.17
Wholesale and retail trade-repair of motor vehicles and personal household goods	-0.32	-0.26	-0.34	-0.36	-0.37	-0.34
Manufacture of food, beverage, and tobacco products	-0.40	-0.40	-0.39	-0.43	-0.44	-0.45
Real estate, renting	-0.50	-0.67	-0.96	-0.37	-0.67	-0.23
Other social activity	-0.55	-0.52	-0.56	-0.62	-0.55	-0.51
Joint Significance of industry effects (p- values)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)

Notes: (1) The coefficients are expressed as deviation from employment weighted average industry effect. (2) Industries are ranked according to the magnitude of industry effects at the mean (given by OLS).

Table 6: Decomposition of Wage Changes over time

Percentile	Marginal-1	Marginal-0	Observed Difference	Coefficient effect	Covariate effect	Residual
Wage differential between 1993-1983						
	1993	**1983**				
Mean*	5.180	4.850	0.330	0.240	0.090	0.000
10^{th}	4.160	3.940	0.210	0.170	0.040	0.000
				0.158; 0.183	0.027; 0.053	
25^{th}	4.740	4.490	0.250	0.220	0.070	-0.040
				0.213; 0.229	0.060; 0.077	
50^{th}	5.300	4.930	0.370	0.280	0.090	0.000
				0.273; 0.285	0.088; 0.101	
75^{th}	5.770	5.290	0.470	0.330	0.120	0.030
				0.319; 0.331	0.116; 0.129	
90^{th}	6.130	5.650	0.480	0.360	0.130	0.000
				0.348; 0.364	0.119; 0.136	
90^{th} - 10^{th}	1.977	1.706	0.271	0.186	0.082	0.003
90^{th} - 50^{th}	0.836	0.722	0.114	0.077	0.033	0.004
50^{th} - 10^{th}	1.141	0.984	0.156	0.109	0.049	-0.001
Wage differential between 2004-1993						
	2004	**1993**				
Mean*	5.380	5.180	0.200	0.250	-0.060	0.000
10^{th}	4.210	4.160	0.060	0.120	-0.100	0.040
				0.104; 0.129	-0.108; -0.085	
25^{th}	4.740	4.740	0.000	0.090	-0.090	0.000
				0.085; 0.102	-0.100; -0.081	
50^{th}	5.360	5.300	0.070	0.170	-0.070	-0.040
				0.165; 0.179	-0.078; -0.060	
75^{th}	6.080	5.770	0.320	0.300	-0.030	0.050
				0.288; 0.304	-0.038; -0.020	
90^{th}	6.560	6.130	0.430	0.400	0.000	0.030
				0.386; 0.403	-0.007; 0.014	
90^{th} - 10^{th}	2.347	1.977	0.370	0.278	0.101	-0.008
90^{th} - 50^{th}	1.197	0.836	0.361	0.223	0.074	0.064
50^{th} - 10^{th}	1.150	1.141	0.010	0.055	0.027	-0.072

Continued…………………

Table 6. ________ (continued)

Percentile	Marginal-1	Marginal-0	Observed Difference	Coefficient effect	Covariate effect	Residual
			Wage differential between 2004-1983			
	2004	**1983**				
Mean*	5.380	4.850	0.520	0.460	0.060	0.000
10^{th}	4.210	3.940	0.270	0.270	-0.040	0.030
				0.262; 0.285	-0.048; -0.027	
25^{th}	4.740	4.490	0.250	0.300	-0.010	-0.040
				0.294; 0.311	-0.020; -0.002	
50^{th}	5.360	4.930	0.440	0.430	0.040	-0.040
				0.426; 0.440	0.036; 0.053	
75^{th}	6.080	5.290	0.790	0.590	0.120	0.070
				0.588; 0.602	0.112; 0.129	
90^{th}	6.560	5.650	0.910	0.720	0.160	0.030
				0.715; 0.733	0.147; 0.169	
90^{th} - 10^{th}	2.347	1.706	0.641	0.452	0.195	-0.005
90^{th} - 50^{th}	1.197	0.722	0.475	0.292	0.114	0.068
50^{th} - 10^{th}	1.150	0.984	0.166	0.160	0.080	-0.074

Note: (1) * Oaxaca-Blinder decomposition.

(2) The second entry is 95% confidence interval for the change. The confidence bounds are the quantiles 2.5% and 97.5% of the distribution of the relevant statistic obtained by bootstrap with 1000 replications.

(3) Marginal-1 and Marginal-0 refers observed marginal wage distributions in year 1 and year 0.

(4) Observed difference is the difference between observed marginal distributions of year 1 and year 0 at each quantile, i.e., α {f(w(1))} – α {f(w(0))}. Year 1 refers the current year and year 0 refers previous year.

(5) Coefficient effect is difference between the counterfactual density (with prices of year1 and covariates of year 0) and the fitted density for year 0, i.e., α{f *(w(1);x(0))} – α{f*(w(0))}.

(6) Covariate effect is difference between fitted density of year 1 and counterfactual density with prices of year1 and covariates of year zero, i.e., α{f*(w(1))} – α{f*(w(1);x(0))}.

(7) Residual is the difference of difference in actual empirical densities and difference in fitted densities at each quantile, i.e, {α(f (w(1))) - α(f(w(0)))} – {α(f*(w(1))) - α(f*(w (0)))}.

See text for further details.

Table 7: JMP Decomposition of Wage changes

	Total Change	Observed Quantities	Observed Prices	Unobservable
Panel I : Wage differential between 1993 and 1983				
10th	0.213	0.023	0.224	-0.034
25th	0.247	0.009	0.212	0.026
50th	0.369	0.077	0.238	0.054
75th	0.474	0.159	0.248	0.067
90th	0.483	0.157	0.248	0.079
90th - 10th	0.271	0.134	0.024	0.113
90th – 50th	0.114	0.080	0.010	0.025
50th – 10th	0.156	0.055	0.014	0.088
Panel II : Wage differential between 2004 and 1993				
10th	0.057	-0.043	0.148	-0.048
25th	0.005	-0.093	0.166	-0.069
50th	0.066	-0.116	0.226	-0.043
75th	0.315	0.005	0.328	-0.017
90th	0.427	0.004	0.402	0.021
90th - 10th	0.370	0.047	0.254	0.069
90th – 50th	0.361	0.120	0.177	0.064
50th – 10th	0.010	-0.073	0.078	0.005

Note: JMP refers to Juhn, Murphy and Pierce (1993)

Appendix

Table A1: OLS and Quantile Regression, 1983

	OLS	q10	q25	q50	q75	q90
constant	2.892***	1.880***	2.416***	3.167***	3.625***	3.900***
primary	0.143***	0.111***	0.136***	0.130***	0.122***	0.092***
middle	0.237***	0.234***	0.209***	0.200***	0.210***	0.188***
secondary	0.529***	0.540***	0.483***	0.482***	0.486***	0.487***
tertiary	0.851***	0.816***	0.777***	0.768***	0.824***	0.891***
age	0.065***	0.078***	0.073***	0.056***	0.050***	0.050***
agesq	-0.001***	-0.001***	-0.001***	-0.001***	-0.000***	-0.000***
female	-0.333***	-0.569***	-0.349***	-0.232***	-0.228***	-0.186***
sc	-0.078***	-0.080**	-0.089***	-0.072***	-0.067***	-0.090***
st	-0.059**	-0.002	-0.011	-0.038	-0.081*	-0.087**
muslim	0.000	0.015	0.017	-0.012	0.012	-0.027
married	0.108***	0.177***	0.142***	0.102***	0.070***	0.056***

Note: All the models include state and industry dummies. * p<0.05; ** p<0.01; *** p<0.001

Table A2: OLS and Quantile Regression, 1993-94

	OLS	q10	q25	q50	q75	q90
constant	3.182***	2.236***	2.737***	3.259***	3.884***	4.240***
primary	0.099***	0.112***	0.130***	0.065***	0.036	0.088***
middle	0.210***	0.183***	0.185***	0.185***	0.184***	0.189***
secondary	0.475***	0.417***	0.448***	0.443***	0.455***	0.475***
tertiary	0.817***	0.747***	0.807***	0.811***	0.816***	0.864***
age	0.058***	0.068***	0.063***	0.057***	0.044***	0.040***
agesq	-0.001***	-0.001***	-0.001***	-0.000***	-0.000***	-0.000***
female	-0.341***	-0.720***	-0.473***	-0.235***	-0.167***	-0.137***
sc	-0.091***	-0.071**	-0.054**	-0.063***	-0.073***	-0.078***
st	-0.026	-0.028	-0.022	-0.027	-0.017	-0.024
muslim	-0.050**	-0.087**	-0.051**	-0.046**	-0.051**	-0.057**
married	0.118***	0.187***	0.157***	0.106***	0.102***	0.059***

Note: All the models include state and industry dummies. * p<0.05; ** p<0.01; *** p<0.001

Table A3: OLS and Quantile Regression, 2004-05

	OLS	q10	q25	q50	q75	q90
constant	3.379***	2.568***	3.208***	3.496***	4.127***	4.455***
primary	0.151***	0.190***	0.148***	0.109***	0.144***	0.151***
middle	0.232***	0.238***	0.218***	0.211***	0.225***	0.198***
secondary	0.546***	0.424***	0.453***	0.503***	0.562***	0.595***
tertiary	1.013***	0.903***	0.932***	0.953***	0.978***	1.046***
age	0.051***	0.060***	0.041***	0.044***	0.035***	0.041***
agesq	-0.000***	-0.001***	-0.000***	-0.000***	-0.000**	-0.000***
female	-0.364***	-0.558***	-0.491***	-0.345***	-0.229***	-0.189***
sc	-0.118***	-0.122***	-0.114***	-0.106***	-0.137***	-0.118***
st	-0.020	-0.006	-0.024	-0.032	-0.100***	-0.051
muslim	-0.074***	-0.069*	-0.073***	-0.032*	-0.063**	-0.102***
married	0.124***	0.149***	0.170***	0.107***	0.084***	0.075***

Note: All the models include state and industry dummies. * p<0.05; ** p<0.01; *** p<0.001