Occupational Tasks and Changes in the Wage Structure

Sergio Firpo — EESP-FGV
Nicole Fortin — UBC
Thomas Lemieux — UBC and NBER

Revised, July 2010

Abstract

This paper looks at the contribution of occupations to changes in the distribution of wages. We first present a simple model where wages in each occupation are set on the basis of a Roy-type linear skill pricing model. We argue that this simple model provides a general way of capturing changes in wages induced by factors like technological change and offshoring. Using Current Population Survey (CPS) for the years 1983-85 and 2000-02, we explore the implications of this model and find that it characterizes well the observed changes in the wage distribution. We then explicitly quantify the contribution of occupational factors, as summarized by the task content of occupations, to changes in wage inequality relative to other explanations such as de-unionization and changes in the returns to education. We do so using a decomposition based on the recentered influence function regression approach of Firpo, Fortin, and Lemieux (2009). The results indicate that technological change and offshorability are two important explanations for the observed changes in the distribution of wages, although other factors like de-unionization and increasing returns to education play a substantial role too.

1 Introduction

Most studies on changes in inequality and the wage structure have focused on explanations based on changes in the returns to traditional measures of skills like education and experience (e.g. Katz and Murphy, 1992) or institutions (e.g. DiNardo, Fortin, and Lemieux, 1996). The role of industrial change due to de-industrialisation and foreign competition was also explored in some of the early studies such as Murphy and Welch (1991), Bound and Johnson (1992), and Freeman (1995). Until recently, however, little attention has been paid to the potential role of occupations in changes in wage inequality.

This situation has changed over the last five years for a number of reasons. First, Autor, Levy, and Murnane (2003), Goos and Manning (2007), and Autor, Katz and Kearney (2006) have proposed a new explanation for changes in wage inequality based on a more "nuanced" view of skill-biased technological change. The idea is that the introduction of computer and information technologies has not simply depressed the relative demand for less skilled workers, as it was assumed in early studies such as Berman, Bound, and Griliches (1994). Rather, what computer and information technologies have done is to depress the return to "routine" tasks that can now be executed by computer technologies, irrespective of whether they require skilled or unskilled labor in the first place. Autor, Katz and Kearney (2006) and Goos and Manning (2007) argue that this nuanced view of technological change can help account for the polarization of wages that has been observed since the late 1980s. Under this type of technological change, it is plausible that moderately skilled workers who used to perform routine tasks experienced a decline in relative wages during this period. To the extent that these workers are around the middle of the skill distribution, technological change could explain why wages in the middle of the distribution fell more than those at the bottom and top end of the distribution.[1]

This more nuanced view of technological change puts occupations at the forefront of the inequality debate since the task content of work (routine nature of the job, cognitive skills required, etc.) is typically measured at the occupational level.[2] Occupations are, therefore, a key empirical channel through which we can assess how technological change affects the wage structure. An important empirical implication of this more nuanced view

[1] Acemoglu and Autor (2010) develop a formal model to show how this could happen in a model with three skill levels (high, middle, and low).

[2] Most studies have either used data from the Dictionary of Occupation Titles (DOT) or the more recent Occupational Information Network (O*NET) to get information about the task content of jobs. Since jobs are defined on the basis of a detailed occupational classification, this naturally lead to an analysis at the occupational level.

of technological change that we discuss below is that changes in the wage structure within and between occupations should be systematically related the type of tasks performed in these occupations.

A second reason for looking at the contribution of occupations in changes in the wage structure is offshoring. Traditional explanations for the role of international trade in changes in inequality have focused on the role of trade in final products that are defined at the industry level. More recently, however, authors such as Feenstra and Hanson (2003) have argued that trade in intermediate inputs was a more promising explanation for changes in wage inequality than trade in final goods and services. For instance, a U.S. multinational can hire computer programmers (middle skilled) in India to update and debug a software product. This lowers the relative demand for that particular occupation, computer programmers, in the United States, which then depresses their wages. The work performed, say overnight, by the offshored programmers, can enhance the productivity of computer software engineers and developers and contribute to wage increases at the other end of the skill spectrum. As in the case of technological change, occupations are the key channel through which offshoring can contribute to changes in wage inequality. The role of offshoring in changes in the wage structure can, thus, be assessed by contrasting the evolution of wages and employment in occupations that are potentially offshorable (e.g. Blinder (2007) and Jensen and Kletzer (2007)) relative to those which are not offshorable.

A third reason for looking at the contribution of occupations in changes in wage inequality is the stunning growth in wages at the very top of the distribution. For instance, Gabaix and Landier (2008) link the growth in CEO pay to an increase in the return to talent in this particular occupation. Kaplan and Rauh (2009) show that workers in a few highly paid occupations in the financial sector also account for a large share of the growth in wages at the very top end of the distribution.

But although occupations now feature prominently as a possible explanation for recent changes in wage inequality, the role of occupations in these changes has not been systemically investigated yet. Some studies do suggest an important role for an occupational based explanation. Goos and Manning (2007) show that the composition effect linked to changes in the distribution of occupations accounts for a substantial part of the increase in inequality in the United Kingdom. Autor, Katz and Kearney (2006) provide evidence that, consistent with a "nuanced" view of technological change, the share of employment in occupations in the middle of the wage distribution has declined over time. Autor and Acemoglu (2010) explore this point in more detail and also show evidence that changes in inter-occupation wage differentials are an important factor in the growth in the variance of U.S. wages since 1980. While these findings suggest a potentially important role for occupations, it remains to be seen how much of the total change in the distribution of wages occupations can precisely account for.

The goal of this paper is to fill this gap by systematically investigating the contribution of occupations to changes in the distribution of wages. We do so by first presenting a conceptual model of the labor market where productive skills are rewarded differently in different occupations, as in a standard Roy model. We argue that this simple model provides a general way of capturing changes in the wage structure induced by factors like technological change and offshorability. Using Current Population Survey (CPS) for the years 1983-85 and 2000-02, we then show evidence that the level and dispersion of wages across occupations have changed substantially over time, and that these changes are linked to the task content of occupations. We measure the task content of occupations using data from the O*NET, and create five indexes of tasks that arguably capture the potential effect of technological change and offshorability on occupational wages. We find that task content measures explain well (at least half of the observed variation) the changes in both the level and the dispersion of wages across occupations. This evidence suggests that changes in the way wages are set across occupations is a promising way of accounting for the U-shaped feature of changes in the wage distribution (Autor, Katz, and Kearney, 2006).

In the last part of the paper, we explicitly quantify the contribution of occupations, as summarized by the task content of jobs, in overall changes in the distribution of wages. We do so using a decomposition method based on the recentered influence function regression approach of Firpo, Fortin, and Lemieux (2009, 2010). This approach enables us to compute the contribution of occupational tasks compared to other explanations such as de-unionization and changes in the labor market wide returns to general skills, such as labor market experience and education. The results indicate that technological change and offshoring are two among a variety of other factors that can account for the observed changes in the distribution of wages.

The paper is organized as follows. In Section 2, we present a Roy model where the returns to a variety of skills can be occupation-specific. This model provides a rationale for looking at the contribution of changes in the occupational wage structure in overall changes in inequality, and helps connect the task content of occupations with wage setting in these occupations. In Section 3 we introduce measures of task content computed from the O*NET data, and explain how we link those to the concepts of technological change and offshorability. Section 4 shows changes in the level and dispersion of wages across occupations, and looks at how these changes are connected to our measures of the task content of jobs. We present the results of our decomposition based on recentered influence function regressions in Section 5, and conclude in Section 6.

2 Wage Setting in Occupations

2.1 Roy Model of Wage Setting

Most of the wage inequality literature follows a traditional Mincerian approach where wages are solely determined on the basis of (observed and unobserved) skills. Equilibrium skill prices depend on supply and demand factors that shape the evolution of the wage structure over time. Underlying changes in demand linked to technological change and offshoring can certainly have an impact on the allocation of labor across industry and occupations, but ultimately wage changes are only linked to changes in the pricing of skills. Acemoglu and Autor (2010) refer to this model as the "canonical model" that has been used in many influential studies such as Katz and Murphy (1992), for example.

There is increasing evidence, however, that the canonical model does not provide a satisfactory explanation for several important changes in the wage structure that have been observed over the last few decades. This is discussed in detail in Acemoglu and Autor (2010) who mention, among other things, two important shortcomings of the canonical model that are particularly important in the context of this paper. First, the canonical model does not easily account for differential changes in inequality in different parts of the distribution that have been observed since the late 1980s (the "polarization" of the wage distribution). Second, the model does not provide insight on the role of occupations in changes in the wage structure because it does not draw any distinction between "skills" and "tasks". Acemoglu and Autor (2010) address these shortcomings by proposing a Ricardian model of the labor market that incorporates a clear distinction between skills and tasks. This model goes a long way towards explaining the recent changes in the wage structure that are hard to account for using the canonical model.

Relative to Acemoglu and Autor (2010), we go one step further here by allowing wages to vary across occupations *conditional* on the skills of workers, as in a standard Roy model. In Acemoglu and Autor's Ricardian model, workers with different levels of skills are systematically allocated to different occupations on the basis of comparative advantage. But, critically, the law of one price holds within each skill group in the sense that wages are equalized across occupations, conditional on skill.

Unlike Acemoglu and Autor, we do not develop a full model of the labor market showing how skills are allocated to occupations (i.e. tasks), and how wages across skills and tasks are set in equilibrium. But our approach that allows wages to vary across occupations, conditional on tasks, follows a long tradition in labor economics. On the theoretical side, Rosen (1983) pointed out that Welch (1969)'s model where wages solely depend on the bundle of skills supplied by each worker to the labor market was unlikely to hold when workers are allocated to a different tasks or occupations. In Welch (1969), the wage w_{it} of worker i at time t is set as follows:

$$w_{it} = \theta_t + \sum_{k=1}^{K} r_{kt} S_{ik} + u_{it}, \tag{1}$$

where S_{ik} (for $k = 1, ..., K$) is each skill component k embodied in worker i, and u_{it} is an idiosyncratic error term. The r_{kt}'s are the returns (or "prices") to each skill component k, while θ_t is a base payment that a worker receives regardless of her skills.

Now consider what happens when workers have the choice between several occupations that have different production functions (or skill requirements). Following Rosen (1983), Acemoglu and Autor (2010) and the approach used in the canonical model, assume that the production function in an occupation simply depends on the sum of skills supplied by all workers in the occupation. In other works, for each skill k, the aggregate supply of skill in the occupation is simply the sum of S_{ik} over all workers i in the occupation. In his comment on Welch (1969), Rosen (1983) shows that returns to skill will only be equalized across occupations if there is sufficient heterogeneity in skill mix across workers to accommodate the large differences in skill requirements across occupations.

To take a simple example, consider two occupations, mathematicians and movers, and two skills, cognitive and physical strength. Clearly, cognitive skills are a particularly productive skill for mathematicians, while physical strength is essential for movers. Say, for instance, that the marginal product of cognitive and physical skills will only be equalized across these two occupations if the cognitive/physical skill ratio is 10 for mathematicians, and 0.1 for movers. Because workers move into an occupation with their own bundle of skills, marginal products will only be equalized if the average ratio of cognitive to physical skills is 100 times larger for mathematicians than for movers. Although people who choose to be mathematicians certainly tend to have a high ratio of cognitive to physical skills, it is very unlikely that people are heterogenous to the point where it is possible to accommodate the skill ratios required in each occupation. Therefore, there will be an oversupply of physical skills among mathematicians that will drive the return to physical skills to almost zero in that occupation. Likewise, there will be an oversupply of cognitive skills among movers that will drive the return to this skill to close to zero. As a result, the return to skill will not be equalized across these two occupations.

The key problem here is that each worker comes with a bundle of skills to be used in a single task or occupation. If skills could be unbundled and efficiently allocated across occupations, returns to skill would all get equalized across occupations, as in Welch (1969). Heckman and Sheinkman (1987) test and resoundly reject the unbundling hypotheses by showing that wages systematically differ across sectors even after controlling for observed and unobserved skills. Gibbons et al. (2005) reach a similar conclusion when looking at both industry or occupation wage differentials.

In other words, there is a wide range of empirical evidence in support of the Roy model of wage determination and self selection where skills are rewarded differently in different occupations, which leads to a systematic sorting of workers into these different occupations. Given the strong theoretical and empirical reasons why wages and returns to skill may not get equalized across occupations, in this paper we explore the consequences on the overall wage structure of differences and changes in wage setting across occupations.[3]

Generalizing equation (1) to the case where returns to skill vary across occupations k (for $k = 1, .., K$) yields the following wage setting equation:

$$w_{ijt} = \theta_{jt} + \sum_{k=1}^{K} r_{jkt} S_{ik} + u_{ijt}, \tag{2}$$

where w_{ijt} is now the wage of individual i in occupation j at time t, the r_{jkt}'s are the returns (or "prices") to each skill component k in occupation j, and θ_{jt} is a base payment that a worker receives in occupation j regardless of her skills.

This wage-setting model is general enough to capture the impact of factors such as technological change or offshoring on wages. For instance, consider the return to manual dexterity. Prior to the introduction of sophisticated robots or other computer technologies, manual dexterity was a highly valued skill in some particular occupations (e.g. precision workers) but not in others (e.g. sales clerk). When routine manual tasks start getting replaced by automated machines or robots, this depresses the return to

[3] A numbers of other reasons such as compensating wage differentials, adjustment costs, or occupation-specific human capital could also be invoked for explaining why wages fail to equalize across occupations, conditional on skill. We focus on Rosen (1983)'s model instead as it provides a rationale for why the return to skill, and not just the level of wages, differs across occupations. This plays a central role when looking at the contribution of occupational wage setting in the overall distribution of wages.

manual dexterity in occupations where these returns were previously high, but not in others where manual dexterity was not a job requirement.

Similarly, returns to social or communication skills are presumably high in occupations where face-to-face meetings with customers are important (e.g. sale representatives). In occupations where face-to-face meetings are not essential (e.g. phone operators), however, the returns to these skills will go down as firms are now able to offshore a lot of this work. The general point here is that the impact of technological change and offshoring can be captured in the above model by changes in the skill pricing parameters r_{jkt}.

2.2 Empirical approach

Ideally, we would like to estimate the skill pricing parameters r_{jkt} using repeated cross sections from a large data set containing detailed information on wages, skills, and occupations. We could then look at the contribution of changes in occupational wage setting in overall changes in the wage structure by computing counterfactual distributions based on alternative measures of r_{jkt} (and θ_{jt}). Unfortunately, no such data set exists. As a result, we instead derive some indirect predictions from our Roy-type wage setting model (2) to look at the contribution of occupational wage setting in changes in the wage structure.

A first general prediction of the model is that if r_{jkt} changes differently in different occupations, this should have an impact on both the between- and within-occupation dimensions of wage inequality. The simple intuition for this prediction is that if the return to a skill heavily used in one occupation goes up, the wage gap between that occupation and others will increase (between-occupation dimension), and so will the wage dispersion within the occupation (within-occupation dimension). One first test of whether changes in occupation-specific skill prices contribute to changes in inequality is, therefore, to see whether there are significant differences in the changes in both the level and the dispersion of wages across occupations.

A second prediction is that changes in both the level and dispersion of wages in occupations should be systematically related to the task content of occupations. For example, in the O*NET data we discuss in the next section, economists get a high score on the task "Analyzing Data or Information". Presumably, having good cognitive skills are quite important for performing these types of tasks. In a Roy model, we expect the return to cognitive skills to be high among economists, so that people with high values of these skills sort into that particular occupation. To the extent that the introduction of better computers increases the marginal product of cognitive skills among economists, we expect both the level and the dispersion of wages (gap between economists with more and less cognitive skills) to increase for economists, or other occupations getting a high score on "Analyzing Data or Information".

To recap, although wages depend solely on skills and occupation-specific returns to skill in equation (2), the task content of occupations should be a useful predictor for changes in both the level and dispersion of wages across occupations. Another advantage of using the task content of occupations is that it reduces the dimensionality of the estimation problem by reducing a large set of occupation dummies into a more limited number of tasks performed in these occupations. We now explain the empirical approach we use to look at these two predictions.

2.2.1 Summarizing changes in the level and dispersion wages using linear regressions for wage quantiles

If the only distributional statistic of interest was the variance, we could compute the contribution of occupations to the overall variance by simply looking at the mean and variance of wages in each occupation, and plugging those into the standard analysis-of-variance formula. Looking at the variance fails to capture, however, the polarization of the wage distribution that has occurred since the late 1980s. As a result, we need an alternative way of summarizing changes in the wage distribution for each occupation that is yet flexible enough to allow for different changes in different parts of the distribution. We do so by estimating linear regression models for the changes in wages at different quantiles of the wage distribution for each occupation. As we now explain in more detail, the intercept and the slope from these regressions are the two summary statistics we use to characterizes the changes in the wage distribution for each occupation.

It is useful to first consider how this approach could be implemented if we had panel data on individuals who stay in the same occupation over time. We could then look at how the wage of worker i changes in response to changes in the skill pricing parameters, Δr_{jk}:

$$\Delta w_{ij} = \Delta\theta_j + \sum_{k=1}^{K} \Delta r_{jk} S_{ik} + \Delta u_{ijt}.$$

The wage change, Δw_{ij}, can be linked to the wage in the base period ($t = 0$) using a simple linear regression equation

$$\Delta w_{ij} = \widetilde{a}_j + \widetilde{b}_j w_{ij0} + e_{ij}.$$

Under the simplifying assumption that the different skill components S_{ik} are uncorrelated, the slope parameter of the regression, $\widetilde{b}_j$, can be written as:

$$\widetilde{b}_j = \frac{cov(\Delta w_{ij}, w_{ij0})}{var(w_{ij0})} = \frac{\sum_{k=1}^{K} (r_{jk0} \Delta r_{jk}) \cdot \sigma_{kj}^2}{\sum_{k=1}^{K} r_{jk0}^2 \cdot \sigma_{kj}^2 + \sigma_{uj0}^2}, \tag{3}$$

where σ_{kj}^2 is the variance of the skill component S_{ik} for workers in occupation k, and where σ_{ujt}^2 is the variance of the idiosyncratic error term u_{ijt}.

Even when the r_{jkt}'s cannot be estimated for lack of precise measures of the skill components S_{ik}, it is still possible to learn something about changes in the r_{jkt}'s from the estimates of the slope coefficients b_j. While the denominator in equation 3 (a variance) is always positive, the sign of the numerator depends on the correlation between returns to skills in the base period (r_{jk0}), and change in the return to skill (Δr_{jk}). For example, in manual occupations where the return to manual dexterity used to be large ($r_{jk0} >> 0$) but declined substantially ($\Delta r_{jk} << 0$), we expect the slope coefficient $\widetilde{b}_j$ to be negative. By contrast, in some scientific or professional occupations where the return to cognitive skills is high ($r_{jk0} >> 0$) and does increase over time ($\Delta r_{jk} >> 0$), we expect the slope coefficient $\widetilde{b}_j$ to be positive.

In the empirical analysis presented below, we rely on large repeated cross-sections of the CPS instead of panel data. While it is not feasible to directly estimate $\widetilde{b}_j$ in that setting, we can still estimate a closely related parameter b_j using percentiles of the within-occupation distribution of wages.

To fix ideas, let's further simplify the model by assuming that both the skill components, S_{ik} (for $k = 1, ..., K$), and the idiosyncratic error term, u_{ijt}, follow a normal distribution. It follows that wages are themselves normally distributed, and the q^{th} percentile of the distribution of w_{ijt}, w_{jt}^q, is given by

$$w_{jt}^q = \overline{w}_{jt} + \sigma_{jt} \Phi^{-1}(q), \tag{4}$$

where $\Phi(\cdot)$ is the standard normal distribution function, $\overline{w}_{jt}$ is the mean of wages in occupation j at time t, and σ_{jt} is its standard deviation, where

$$\sigma_{jt}^2 = \sum_{k=1}^{K} r_{jkt}^2 \cdot \sigma_{kj}^2 + \sigma_{ujt}^2.$$

Now consider a regression of Δw_j^q on w_{j0}^q:[4]

$$\Delta w_j^q = a_j + b_j w_{j0}^q + e_j^q. \tag{5}$$

The slope parameter, b_j, is now given by

$$b_j = \frac{cov(\Delta w_j^q, w_{j0}^q)}{var(w_{j0}^q)} = \frac{(\Delta\sigma_j \cdot \sigma_{j0}) \cdot var(\Phi^{-1}(q))}{\sigma_{j0}^2 \cdot var(\Phi^{-1}(q))} = \frac{\Delta\sigma_j}{\sigma_{j0}}. \tag{6}$$

Using the linear approximations $\Delta\sigma_j \approx \Delta\sigma_j^2/2\sigma_{j0}$ and $\Delta r_{jk}^2 \approx 2r_{jk0}\Delta r_{jk}$ yields

$$b_j \approx \frac{\sum_{k=1}^{K} (r_{jk0}\Delta r_{jk}) \cdot \sigma_{kj}^2}{\sigma_{j0}^2} + \frac{\Delta\sigma_{uj}^2}{2\sigma_{j0}^2}. \tag{7}$$

The first term in equation (7) is similar to the slope coefficient obtained earlier in (3) and has, therefore, a similar interpretation. The second term reflects the fact that an increase in the variance of the idiosyncratic error term widens the wage distribution, which results in a positive relationship between changes in wages and base wage levels.

Using the fact that $E(w_{jt}^q) = \overline{w}_{jt}$ (expectation taken over q, for $q = 0,..,1$), the intercept in the regression model, a_j, can be written as:

$$a_j = \Delta\overline{w}_j - \frac{\Delta\sigma_j}{\sigma_{j0}}\overline{w}_{j0}, \tag{8}$$

where

$$\Delta\overline{w}_j = \Delta\theta_j + \sum_{k=1}^{K} \Delta r_{jk}\overline{S}_{jk}. \tag{9}$$

Without loss of generality, we can normalize the base period wage in each occupation

[4] Note that under the normality assumption, the error term e_j^q is equal to zero. We introduce the error term in the equation, nonetheless, to later allow for a more general case where the normality assumption fails.

to have a mean zero. The intercept can then be written as:

$$a_j = \Delta\theta_j + \sum_{k=1}^{K} \Delta r_{jk} \overline{S}_{jk}. \tag{10}$$

Like the slope parameter b_j, the intercept a_j depends on changes in the return to skill components, Δr_{jk}. The intercept also depends on $\Delta\theta_j$, which reflects changes in occupational wage differentials unrelated to skills. This could reflect occupational rents, compensating wage differentials, etc. Under the strong assumption that skills S_{ik} and the error term u_{ijt} are normally distributed, the regression model in equation (5) fully describes the relationship between the base wage and the change in wage at each percentile q of the within-occupation wage distribution.

This suggests a simple way of assessing the contribution of changes in the occupational wage structure to changes in the distribution of wages. In a first step, we can estimate equation (5) separately for each occupation (or in a pooled regression with interactions) and see to what extent the simple linear model helps explain the observed changes in wages. We can then run "second step" regressions of the estimated a_j and b_j on measures of task content of work that correlates with the r's and with the change in the r's at the occupational level.

While the normality assumption is convenient for illustrating the basic predictions of the linear skill pricing model, it is also restrictive. As is well known, the normal distribution is fully characterized by its location ($\overline{w}_{jt}$ above) and scale parameter (σ_{jt} above). This can be generalized to the case where the wage distribution is not normal, but only the location and scale changes over time. Relative to equation (4), this means we can replace $\Phi^{-1}(q)$ by a more general and occupation-specific inverse probability function $F_j^{-1}(q)$. Equation (4) is then replaced by

$$w_{jt}^q = \overline{w}_{jt} + \sigma_{jt} F_j^{-1}(q). \tag{11}$$

We can then get the same regression equation since

$$\Delta w_j^q = \Delta\overline{w}_j + \Delta\sigma_j F_j^{-1}(q). \tag{12}$$

Solving for $F_j^{-1}(q)$ in equation (11) at $t = 0$, and substituting into equation (12) yields

$$\Delta w_j^q = \Delta\overline{w}_j - \frac{\Delta\sigma_j}{\sigma_{j0}} \overline{w}_{j0} + \frac{\Delta\sigma_j}{\sigma_{j0}} w_{j0}^q, \tag{13}$$

which is identical to equation (5) since $a_j = \Delta\overline{w}_j - \frac{\Delta\sigma_j}{\sigma_{j0}}\overline{w}_{j0}$ and $b_j = \frac{\Delta\sigma_j}{\sigma_{j0}}$.

In general, however, changes in the returns to skill r_{jkt} are expected to change the shape of the wage distribution above and beyond the scale and location; $F_j^{-1}(q)$ is no longer a constant over time. As a result, equation (12) becomes

$$\begin{aligned} \Delta w_j^q &= \Delta\overline{w}_j + \left(\sigma_{j1}F_{j1}^{-1}(q) - \sigma_{j0}F_{j0}^{-1}(q)\right) && (14) \\ &= \Delta\overline{w}_j + \Delta\sigma_j F_{j0}^{-1}(q) + e_j^q && (15) \end{aligned}$$

where $e_j^q = \sigma_{j1}\left(F_{j1}^{-1}(q) - F_{j0}^{-1}(q)\right)$. Substituting in $F_{j0}^{-1}(q) = (w_{j0}^q - \overline{w}_{j0})/\sigma_{j0}$ and using the definitions of a_j and b_j then yields

$$\Delta w_j^q = a_j + b_j w_{j0}^q + e_j^q. \tag{16}$$

It is generally not possible to find a close form expression for e_j^q. If changes in the $F_{jt}^{-1}(q)$ functions are similar across occupation, however, this will generate a percentile specific component in the error term. For instance, Autor, Katz and Kearney (2008) show that the distribution of wage residuals has become more skewed over time (convexification of the distribution). This is inconsistent with the normality assumption, but can be captured by allowing for a percentile-specific component λ^q in e_j^q:

$$e_j^q = \lambda^q + \varepsilon_j^q. \tag{17}$$

This leads to the main regression equation to be estimated in the first step of the empirical analysis:

$$\Delta w_j^q = a_j + b_j w_{j0}^q + \lambda^q + \varepsilon_j^q. \tag{18}$$

A more economically intuitive interpretation of the percentile-specific error components λ^q is that it represents a generic change in the return to unobservable skills of the type considered by Juhn, Murphy, and Pierce (1993). For example, if unobservable skills in a standard Mincer type regression reflect unmeasured school quality, and that school quality is equally distributed and rewarded in all occupations, then changes in the return to school quality will be captured by the error component λ^q.

2.2.2 Connecting to occupational task measures

In the second step of the analysis, we link the estimated intercepts and slopes (a_j and b_j) to measures of the task content of each occupation. Since technological change and offshoring are the key explanatory variables used in the second step, the next section discusses in detail how we construct five summary measures of occupational tasks. For the time being, define these summary measures of tasks as TC_{jh}, for $h = 1, .., 5$.

The second step regressions are

$$a_j = \gamma_0 + \sum_{h=1}^{5} \gamma_{jh} TC_{jh} + \mu_j, \tag{19}$$

and

$$b_j = \delta_0 + \sum_{h=1}^{5} \delta_{jh} TC_{jh} + \nu_j. \tag{20}$$

There is no direct mapping from the task content measures TC_{jh} to the return to skill parameters, r_{jk}. We expect to see, however, a steeper decline in the relevant r_{jkt}'s in occupations with traditional task requirements that are more easily replaceable by technology or offshore workers. For example, occupations scoring high in terms of the routine aspect of the work performed should experience a sizable decline in both a_j and b_j. Similarly, occupations that involve face-to-face meetings are less likely to be offshored and experience a decline in the a_j or b_j parameters.

3 Occupational Measures of Technological Change and Offshoring Potential

Like many recent papers (Goos and Manning (2007), Goos, Manning and Salomons (2009), Crino (2009)) that study the task content of jobs, and in particular their offshorability, we use the O*NET data to compute our measures of technological change and offshoring potential.[5] Our aim is to produce indexes for all 3-digit occupations available in the CPS, a feat that neither Jensen and Kletzer (2007) nor Blinder (2007) completed.[6] Our construction of an index of potential offshorability follows the pioneering work of Jensen and Kletzer (2007) [JK, thereafter] while incorporating some of the criticisms of Blinder (2007). The main concern of Blinder (2007) is the inability of the objective indexes to take into account two important criteria for non-offshorability: a)

[5] Available from National Center for O*NET Development. We use the O*NET 13.0 which has many updated elements by comparison with the O*NET 10.0 used in Goos, Manning and Salomons (2009).

[6] Blinder (2007) did not compute his index for Category IV occupations (533 occupations out of 817), that are deemed impossible to offshore. Although, Jensen and Kletzer (2007) report their index for 457 occupations, it is not available for many blue-collar occupations (occupations SOC 439199 and up).

that a job needs to be performed at a specific U.S. location, and b) that the job requires face-to-face personal interactions with consumers. We thus pay particular attention to the "face-to-face" and "on-site" categories in the construction of our indexes.

In the spirit of Autor, Levy, and Murnane (2003), who used the Dictionary of Occupational Titles (DOT) to measure the routine vs. non-routine, and cognitive vs. non-cognitive aspects of occupations, JK use the information available in the O*NET, the successor of the DOT, to construct their measure. The O*NET content model organize the job information into a structured system of six major categories: worker characteristics, worker requirements, experience requirements, occupational requirements, labor market characteristics, and occupation-specific information.

Like JK, we focus on the "occupational requirements" of occupations, but we add some "work context" measures to enrich the "generalized work activities" measures. JK consider eleven measures of "generalized work activities", subdivided into five categories: 1) on information content: getting information, processing information, analyzing data or information, documenting/recording information; 2) on internet-enabled: interacting with computers; 3) on face-to-face contact: assisting or caring for others, performing or working directly with the public, establishing or maintaining interpersonal relationships; 4) on the routine or creative nature of work: making decisions and solving problems, thinking creatively; 5) on the "on-site' nature of work: inspecting equipment, structures or material.

We consider five similar categories, "Information Content" and "Automation", thought to be positively related to offshoring, and "Face-to-Face", "On-Site Job", and "Decision-Making", thought to be negatively related to offshoring. Table 1 lists the exact O*NET reference number of the generalized work activities and work context items that make up the five indexes and indicate the elements also used by JK and/or Blinder (2007). Our first category "Information Content" regroups JK categories 1) and 2). It identifies occupations with high information content that are likely to be affected by ICT technologies; they are also likely to be offshored if there are no mitigating factor. Appendix Figure 1 shows that average occupational wages in 2000-02 increase steadily with the information content. Our second category "Automation/Routinization" is constructed using some work context measures to reflect the degree of potential automation of jobs and is similar in spirit to the manual routine index of Autor et al. (2003). The work context elements are: degree of automation, importance of repeating same tasks, structured versus unstructured work (reverse), pace determined by speed of equipment, and spend time making repetitive motions. As shown in Appendix Figure 1, the relationship between our automation index and average occupational wages display an inverse U-shaped left-of-center of the wage distribution. We think of these first two categories as being more closely linked to technological change, thus we called the group "Technology". We agree with Blinder (2007) that there is some degree of overlap with offshorability. Indeed, the information content is a substantial component of JK's offshorability index.

Our three remaining categories "Face-to-Face Contact", "On-Site Job" and "Decision-Making" are meant to capture features of jobs that cannot be offshored. Note, however, that the decision-making features were also used by Autor et al. (2003) to capture the notion of non-routine cognitive tasks. Our "Face-to-Face Contact" measure adds one work activity "coaching and developing others" and one work context "face-to-face discussions" element to JK's face-to-face index. Our "On-Site Job" measure adds four other elements of the JK measure: handling and moving objects, controlling machines and processes, operating vehicles, mechanized devices, or equipment, and repairing and maintaining mechanical equipment and electronic equipment (weight of 0.5 to each of these last two elements). Our "Decision-Making" measure adds one work activity "developing objectives and strategies" and two work context elements, "responsibility for outcomes and results" and "frequency of decision making" to the JK measure. We use the reverse of these measures of non-offshorability to capture "Offshorability". The relationship between these measures and average occupational wages are displayed in Appendix Figure 1. Automation and No-Face-to-Face contact exhibit a similar shape. No-On-Site (Off-Site) is clearly U-shaped, and No-Decision-Making is steadily decreasing with average occupational wages.

For each occupation, O*NET provides information on the "importance" and "level" of required work activity and on the frequency of five categorical levels of work context.[7] We follow Blinder (2007) in arbitrarily assigning a Cobb-Douglas weight of two thirds to "importance" and one third to "level" in using a weighed sum for work activities. For work contexts, we simply multiply the frequency by the value of the level.

Each composite TC_h score for occupation j in category h is, thus, computed as

$$TC_{jh} = \sum_{k=1}^{A_h} I_{jk}^{2/3} L_{jk}^{1/3} + \sum_{l=1}^{C_h} F_{jl} * V_{jl}, \tag{21}$$

[7] For example, the work context element "frequency of decision-making" has five categories: 1) never, 2) once a year or more but not every month, 3) once a month or more but not every week, 4) once a week or more but not every day, and 5) every day. The frequency corresponds to the percentage of workers in an occupation who answer a particular value. As shown in Appendix Table B1, 33 percent of sales manager answer 5) every day, 30 percent of computer programmers and 53 percent of computer engineers answer 4) once a week or more.

where A_h is the number of work activity elements, and C_h the number of work context elements in the category TC_h, $h = 1, \ldots, 5$.

To summarize, we compute five different measures of task content using the O*NET: *i)* the information content of jobs, *ii)* the degree of automation of the job and whether it represents routine tasks, *iii)* the importance of face-to-face contact, *iv)* the need for on-site work, and *v)* the importance of decision making on the job. We use these measures to assess both the impact of technological change and offshorability on changes in wages.

4 Occupation Wage Profiles: Results

In this Section, we present the estimates of the linear regression models for within-occupation quantiles (equation (18)), and then link the estimated slope and intercept parameters to our measures of task content from the O*NET. We refer to these regressions as "occupation wage profiles". The empirical analysis is based on data for men from the 1983-85 and 2000-02 Outgoing Rotation Group (ORG) Supplements of the Current Population Survey. The data files were processed as in Lemieux (2006b) who provides detailed information on the relevant data issues. The wage measure used is an hourly wage measure computed by dividing earnings by hours of work for workers not paid by the hour. For workers paid by the hour, we use a direct measure of the hourly wage rate. CPS weights are used throughout the empirical analysis. By pooling three years of data at each end of the sample period, we obtain relatively large samples both in 1983-85 (274,625 observations) and 2000-02 (252,397 observations).

The choice of years is driven by data consistency issues. First, there is a major change in occupation coding in 2003 when the CPS switches to the 2000 Census occupation classification. This makes it hard to compare detailed occupations from the 1980s or 1990s to those in the post-2002 data. Another data limitation is that union status, another important source of changes in the wage distribution, is only available in the ORG CPS from 1983 on. But despite these data limitations, the 1983-85 to 2000-02 is a good time span to study given the purpose of this paper, since this is when most of the polarization of wages phenomenon documented by Autor, Katz and Kearney (2006) happened.

Indeed, Figure 1 shows that the changes in real wages at each percentile of the wage distribution follow a U-shaped curved, as in Autor, Katz and Kearney (2006). In this figure, we show separately the changes before and after the late 1980s (1988-90), since it is usually believed that polarization is a phenomena of the 1990s. The figure shows that wage changes in the top half of the distribution were quite similar during both time periods. In particular, wages at the very top increased much more than wages in the middle of the distribution, resulting in increased top-end inequality. By contrast, inequality in the lower half of the distribution was stable during the 1980s, but decreased sharply after 1988-90 as wages at the bottom grew substantially more than in the middle of the distribution. So while the U-shaped feature of wage changes only starts to emerge in the 1990s, inequality at the top and bottom were already moving in distinct directions in the second half of the 1980s.

For the sake of simplicity, we focus our analysis on the whole 1983-85 to 2000-02 period. Results for the more recent 1988-90 to 2000-02 period are qualitatively similar, though generally smaller in quantitative terms since the raw changes to be explained are also smaller, at least at the top end of the distribution.

Note that, despite our large samples based on three years of pooled data, we are left with a small number of observations in many occupations when we work at the three-digit occupation level. In the analysis presented in this section, we thus focus on occupations classified at the two-digit level (40 occupations) to have a large enough number of observations in each occupation.[8] This is particularly important given our empirical approach where we run regressions of change in wages on the base-period wage. Sampling error in wages generates a spurious negative relationship between base-level wages and wage changes that can be quite large when wage percentiles are imprecisely estimated.[9]

In principle, we could use a large number of wage percentiles w_{jt}^q in the empirical analysis. But since wage percentiles are strongly correlated for small differences in q, we only extract the nine deciles of the within-occupation wage distribution, i.e. w_{jt}^q for $q = 10, 20, ..., 90$. Finally, all the regression estimates are weighted by the number of observations (weighted using the earnings weight from the CPS) in each occupation.

Figure 2a presents the raw data used in the analysis. The figure plots the 360 observed changes in wages (9 observation for each of the 40 occupations) as a function of the base

[8] Though there is a total of 45 occupations at the two-digit level, we combine five occupations with few observations to similar but larger occupations. Specifically, occupation 43 (farm operators and managers) and 45 (forestry and fishing occupations) are combined with occupation 44 (farm workers and related occupations). Another small occupation (20, sales related occupations) is combined with a larger one (19, sales workers, retail and personal services). Finally two occupations in which very few men work (23, secretaries, stenographers, and typists, and 27, private household service occupations) are combined with two other larger occupations (26, other administrative support, including clerical, and 32, personal services, respectively).

[9] The bias could be adjusted using a measurement-error corrected regression approach, as in Card and Lemieux, 1996, or using an instrumental variables approach.

wages. The most noticeable feature of Figure 1 is that wage changes exhibit the well-known U-shaped pattern documented by Autor, Katz, and Kearney (2006), that we also see in Figure 1. Broadly speaking, the goal of the first part of the empirical analysis is to see whether the simple linear model presented in equation (18) helps explain a substantial part of the variation documented in Figure 2a.

Table 2 shows the estimates from various versions of equation (18). We present two measures of fit for each estimated model. First, we report the adjusted R-square of the model. Note that even if the model in equation (18) was the true wage determination model, the regressions would not explain all of the variation in the data because of the residual sampling error in the estimated wage changes. The average sampling variance of wage changes is 0.0002, which represents about 3 percent of the total variation in wage changes by occupation and decile. This means that one cannot reject the null hypothesis that sampling error is the only source of residual error (i.e. the model is "true") whenever the R-square exceeds 0.97.

The second measure of fit consists of looking at whether the model is able to explain the U-shaped feature of the raw data presented in Figure 2a. As a reference, the estimated coefficient on the quadratic term in the fitted (quadratic) regression reported in Figure 2a is equal to 0.125. For each estimated model, we run a simple regression of the regression residuals on a linear and quadratic term in the base wage to see whether there is any curvature left in the residuals that the model is unable to explain.

One potential concern with this regression approach is that we are not controlling for any standard covariates, which means that we may be overstating the contribution of occupations in changes in the wage structure. For instance, workers with high levels of education tend to work in high wage occupations. This means that changes in the distribution of wages in high wage occupation may simply be reflecting changes in the return to education among highly educated workers. Changes in the distribution of education, or other covariates, may also be confounding the observed changes in occupational wages.

We address these issues by reweighting the distribution of covariates in each occupation at each time period so that it is the same as in the pooled distribution with all occupations and time periods (1983-85 and 2000-02). This involves computing 80 separate logits (40 occupations times two years) to perform a DiNardo, Fortin, and Lemieux (1996) reweighting exercise. The various quantiles of the wage distribution for each occupation are then computed in the reweighted samples. The covariates used in the logit are the same as the explanatory variables used in the influence function regressions of the next Section.[10] The unadjusted models are reported in Panel A of Table 2, while the estimates that adjust for the covariates by reweighting are reported in Panel B. Since the results with and without the adjustment are qualitatively similar, we focus our discussion on the unadjusted estimates reported in Panel A.

As a benchmark, we report in column 1 the estimates from a simple model where the only explanatory variable is the base wage. This model explains essentially none of the variation in the data as the adjusted R-square is only equal to 0.0023. This reflects the fact that running a linear regression on the data reported in Figure 2a essentially yields a flat line. Since the linear regression cannot, by definition, explain any of the curvature of the changes in wages, the curvature parameter in the residuals (0.0125) is exactly the same as in the simple quadratic regression discussed above.

In column 2, we only include the set of occupation dummies (the a_j's) in the regression. The restriction imbedded in this model is that all the wage deciles within a given occupation increase at the same rate, i.e. there is no change in within-occupation wage dispersion. Just including the occupation dummies explain almost two thirds of the raw variation in the data, and half of the curvature. The curvature parameter declines from 0.125 to 0.065 but remains strongly significant.

Column 3 shows that only including decile dummies (the λ^q's) explains none of the variation or curvature in the data. This is a strong result as it indicates that using a common within-occupation change in wage dispersion cannot account for any of the observed change in wages. Interestingly, adding the decile dummies to the occupation dummies (column 4) only marginally improves the fit of the model compared to the model with occupation dummies in column 2. This indicates that within-occupation changes in the wage distribution are highly occupation-specific, and cannot simply be linked to a pervasive increase in returns to skill "à la" Juhn, Murphy and Pierce (1993).

By contrast, the fit of the model improves drastically once we introduce occupation-specific slopes in column 5. The R-square of the model jumps to 0.9310, which is quite close to the critical value for which we cannot reject the null hypothesis that the model is correctly specified, and that all the residual variation is due to sampling error. The curvature parameter now drops to -0.001 and is no longer statistically significant. In other words, we are able to account for all the curvature in the data using occupation-specific slopes. Note also that once the occupation-specific slopes are included, decile

[10] One small difference is that we only use five education categories instead of six in the next section (we combine people with 0-8 and 9-11 years of schooling together). The other covariates are a set of 9 experience dummies, and dummies for race and marital status).

dummies play a more substantial role in the regressions, as evidenced by the drop in the adjusted R-square between column 5 (decile dummies included) and 6 (decile dummies excluded).[11]

The results reported in Panel B where we control for standard covariates are generally similar to those reported in Panel A. In particular, the model with decile dummies and occupation-specific slopes (Figure 5) explains most of the variation in the data and all of the curvature. Note that the R-square is generally lower than in the models where we do not control for covariates. This indicates that the covariates reduce the explanatory power of occupations by relatively more than they reduce the residual variation unexplained by occupational factors. In other words, this reflects the fact that occupational affiliation is strongly correlated with observable skill measures (see, for example, Gibbons et al., 2005).

We next illustrate the fit of the model by plotting all 40 occupation-specific regression curves in Figure 2b. While it is not possible to see what happens in each and every occupation on this graph, there is still a noticeable pattern in the data. The slope for occupations at the bottom end of the distribution tends to be negative. Slopes get flatter in the middle of the distribution, and generally turn positive at the top end of the distribution. In other words, it is clear from the figure that the occupational wage profile generally follow the U-shaped pattern observed in the raw data. This is consistent with the model of Section 2 where the skills that used to be valuable in lower paid occupations are less valuable than they used to be, while the opposite is happening in high-wage occupations.

We explore this hypothesis more formally by estimating the regression models in equations (19) and (20) that link the intercept and slopes of the occupation wage change profiles to the task content of occupations.[12] The results are reported in Table 3. In the first six columns of Table 3, we include all five task measures together in the regressions. Most of the results reported in the Table are based on the models where we do not control for covariates, though we also show estimates based on the reweighted samples in columns 3 and 6. In the last two columns of the table, we show the estimates from

[11] Replacing the decile dummies by a linear function in the "normits" $\Phi^{-1}(q)$ does not substantially affect the fit of the model. One could also include the value of q instead of the normit $\Phi^{-1}(q)$. However, if wages are distributed log normal and the variance within occupation increases by the same in all occupations, this change will be exactly captured by including a linear function in $\Phi^{-1}(q)$. Empirically, a linear function function in $\Phi^{-1}(q)$ indeed fits the data better than a linear function in q.

[12] To be consistent with equation (10), we have recentered the observed wage changes so that the intercept for each occupation corresponds to the predicted change in wage at the median value of the base wage.

regressions in which each task measure is entered separately. To adjust for the possible confounding effect of overall changes in the return to skill, we also report estimates that control for the base (median) wage level in the occupation.

To get a better sense of how these task measures vary across the occupation distribution, consider again Appendix Figure 1, which plots the values of the task index as a function of the average wage in the (3-digit) occupation. The "information content" and "decision making" measures are strongly positively related to wages. Consistent with Autor, Levy and Murnane (2003), the "automation" task follows an inverse U-shaped curve. To the extent that technological change allows firms to replace workers performing these types of tasks with computer driven technologies, we would expect both the intercept and slope of occupations with high degree of automation to decline over time.

But although occupations in the middle of the wage distribution may be most vulnerable to technological change, they also involve relatively more on-site work (e.g. repairmen) and may, therefore, be less vulnerable to offshoring. The last measure of task, face-to-face contact, is not as strongly related to average occupational wages as the other task measures. On the one hand, we expect workers in occupations with a high level of face-to-face contact to do relatively well in the presence of offshoring. On the other hand, since many of these workers may have relatively low formal skills such as education (e.g. retail sales workers), occupations with a high level of face-to-face contact may experience declining relative wages if returns to more general forms of skills increase.

The strongest and most robust result in Table 3 is that occupations with high level of automation experience a relative decline in both the intercept and the slope of their occupational wage profiles. The effect is statistically significant in most specifications reported in Table 3. The other "technology" variable, information content, has a positive and significant effect on both the intercept and the slope, as expected, when included by itself in columns 7 and 8. The effect tends to be weaker, however, in models where other tasks are also controlled for.

The effect of the tasks related to the offshorability of jobs are reported in the last three rows of the table. Note that since "on-site", "face-to-face", and "decision making" are negatively related to the offshorability of jobs, we use the reverse of these tasks in the regression to interpret the coefficients as the impact of offshorability (as opposed to non-offshorability). As a result, we expect the effect of these adjusted tasks to be negative. For instance, the returns to skill in jobs that do not require face-to-face contacts will likely decrease since it is now possible to offshore these types of jobs to another country.

The results reported in Table 3 are mixed. As expected, the effect of no face to face is generally negative. In the case of "no decision making", the estimates are negative and significant when we don't control for other tasks (columns 7 and 8), but are generally insignificant when other tasks are controlled for. Finally, the effect of "no on-site work" is generally positive, which is surprising. One possible explanation is that the O*NET is not well suited for distinguishing whether a worker has to work on "any site" (i.e. an assembly line worker), vs. working on a site in the United States (i.e. a construction worker).

On balance, most of the results reported in Table 3 are consistent with our expectations. More importantly, these tasks explain most of the variation in the slopes, and about half of the variation in the intercepts. This suggests that we can capture most of the effect of occupations on the wage structure using only a handful of task measures, instead of a large number of occupation dummies. The twin advantage of tasks over occupations is that they are a more parsimonious way of summarizing the data, and are more economically interpretable than occupation dummies.

We draw two main conclusions from Table 3. First, as predicted by the linear skill pricing model of Section 2, the measures of task content of jobs tend to have a similar impact on the intercept and the slope of the occupational profiles. Second, tasks account of most of the variation in the slopes and intercepts over occupations, and the estimated effect of tasks are generally consistent with our theoretical expectations. Taken together, this suggests that occupational characteristics as measured by these five tasks can play a substantial role in explaining the U-shaped feature of the raw data illustrated in Figure 1. This suggests that occupational characteristics, as measured by these task content measures, can go a long way towards explaining changes in the wage distribution between 1983-85 and 2000-02. We next examine this issue in more detail in the context of a formal decomposition of changes in the wage distribution.

5 Decomposition: Occupational Characteristics vs. Other Factors

Although the analysis presented in Section 4 helps illustrate the mechanisms through which occupations play a role in changes in the wage structure, it does not precisely quantify the contribution of occupational factors for two reasons. First, the occupational wage profiles estimated above only describe changes in the occupation-specific distribution of wages. One needs to aggregate these wage profiles to quantify the contribution of occupational factors on the overall distribution of wages. Second, while linear specifications are useful for summarizing the link between occupational wage changes and tasks measures, a more flexible approach is required for looking at what happens at various points of the wage distribution. More generally, most of the estimates presented above do not control for other factors such as education, experience, unionization, etc. While some of the estimates (based on reweighted samples) do control for some of these other factors, one cannot readily assess the relative importance on the different explanatory factors from the results reported in Section 4.

In this section, we formally decompose changes in the distribution of wages between 1983-85 and 2000-02 into the contribution of occupational and other factors. We do so using the recentered influence function (RIF) regression approach introduced by Firpo, Fortin, and Lemieux (2009). As is well known, a standard regression can be used to perform a Oaxaca-type decomposition for the mean of a distribution. RIF regressions allow us to perform the same kind of decomposition for any distributional parameter, including percentiles.

Firpo, Fortin, and Lemieux (2007, 2010), explain in detail how to perform these decompositions, and show how to compute the standard errors for each element of the distribution. In this paper, we only provide a quick summary of how the decomposition method works.

5.1 Decomposing Changes in Distributions Using RIF Regressions

In general, any distributional parameter can be written as a functional $\nu(F_Y)$ of the cumulative distribution of wages, $F_Y(Y)$.[13] Examples include wage percentiles, the variance of log wage, the Gini coefficient, etc. The first part of the decomposition consists of dividing the overall change in a given distributional parameter into a composition effect linked to changes in the distribution of the covariates, X, and a wage structure effect that reflects how the conditional distribution of wage $F(Y|X)$ changes over time. In a standard Oaxaca decomposition, the wage structure effect only depends on changes in the conditional mean of wages, $E(Y|X)$. More generally, however, the wage structure effect depends on the whole conditional wage distribution.

It is helpful to discuss the decomposition problem using the potential outcomes frame-

[13] In this section, we denote the wage using Y instead of W to be consistent with Firpo, Fortin, and Lemieux (2007) and the program evaluation literature.

work. We focus on differences in the wage distributions for two time periods, 1 and 0. For a worker i, let Y_{1i} be the wage that would be paid in period 1, and Y_{0i} the wage that would be paid in period 0. Since a given individual i is only observed in one of the two periods, we either observe Y_{1i} or Y_{0i}, but never both. Therefore, for each i we can define the observed wage, Y_i, as $Y_i = Y_{1i} \cdot D_i + Y_{0i} \cdot (1 - D_i)$, where $D_i = 1$ if individual i is observed in period 1, and $D_i = 0$ if individual i is observed in period 0. There is also a vector of covariates $X \in \mathcal{X} \subset \mathbb{R}^K$ that we can observe in both periods.

Consider Δ_O^ν, the overall change over time in the distributional statistic ν. We have

$$\begin{aligned}\Delta_O^\nu &= \nu\left(F_{Y_1|D=1}\right) - \nu\left(F_{Y_0|D=0}\right) \\ &= \underbrace{\nu\left(F_{Y_1|D=1}\right) - \nu\left(F_{Y_0|D=1}\right)}_{\Delta_S^\nu} + \underbrace{\nu\left(F_{Y_0|D=1}\right) - \nu\left(F_{Y_0|D=0}\right)}_{\Delta_X^\nu},\end{aligned}$$

where Δ_S^ν is the wage structure effect, while Δ_X^ν is the composition effect. Key to this decomposition is the counterfactual distributional statistics $\nu\left(F_{Y_0|D=1}\right)$. This represents the distributional statistic that would have prevailed if workers observed in the end period $(D = 1)$ had been paid under the wage structure of period 0.

Estimating that counterfactual distribution is a well known problem. For instance, DiNardo, Fortin and Lemieux (1996) suggest estimating this counterfactual by reweighting the period 0 data to have the same distribution of covariates as in period 1. We follow the same approach here, since Firpo, Fortin and Lemieux (2007) show that reweighting provides a consistent nonparametric estimate of the counterfactual under the ignorability assumption.

However, the main goal of this paper is to separate the contribution of different subsets of covariates to Δ_O^ν , Δ_S^ν , and Δ_X^ν. As is well known, this is easily done in the case of the mean where each component of the above decomposition can be written in terms of the regression coefficients and the mean of the covariates. In that case $\nu\left(F_{Y_1|D=1}\right)$ is just the mean of wages in period 1, and can be written as

$$\nu\left(F_{Y_1|D=1}\right) = E(Y_1|D=1) = E\left[X|D=1\right]^\intercal \beta_1,$$

where β_1 is the vector or coefficient from a standard wage regression in period 1. The contributions of each covariate to the wage structure and composition effect are simply

$$\Delta_S^\mu = E\left[X|D=1\right]^\intercal (\beta_1 - \beta_0),$$

and

$$\Delta_X^{\mu} = (E[X|D=1] - E[X|D=0])^{\intercal}\beta_0.$$

For distributional statistics besides the mean, Firpo, Fortin, and Lemieux (2009) suggest estimating a similar regression where the usual outcome variable, Y, is replaced by the recentered influence function $\mathrm{RIF}(y;\nu)$. The recentering consists of adding back the distributional statistic ν to the influence function $\mathrm{IF}(y;\nu)$:

$$\mathrm{RIF}(y;\nu) = \nu + \mathrm{IF}(y;\nu).$$

Note that in the case of the mean where the influence function is $\mathrm{IF}(y;\mu) = y - \mu$, we have $\mathrm{RIF}(y;\mu) = \mu + (y-\mu) = y$. Since the RIF is simply the outcome variable y, the RIF regression corresponds to a standard wage regression, as in the Oaxaca decomposition.

It is also possible to compute the influence function for any other distributional statistics. Of particular interest is the case of quantiles. The τ-th quantile of the distribution F is defined as the functional, $Q(F,\tau) = \inf\{y|F(y) \geq \tau\}$, or as q_τ for short. Its influence function is:

$$\mathrm{IF}(y;q_\tau) = \frac{\tau - \mathbb{I}\{y \leq q_\tau\}}{f_Y(q_\tau)}. \tag{22}$$

The recentered influence function of the τ^{th} quantile is $\mathrm{RIF}(y;q_\tau) = q_\tau + \mathrm{IF}(y;q_\tau)$. Consider γ^ν, the estimated coefficients from a regression of $\mathrm{RIF}(y;\nu)$ on X. By analogy with the Oaxaca decomposition, the wage structure and composition effects can be written as:

$$\Delta_S^{\nu} = E[X|D=1]^{\intercal}(\gamma_1^{\nu} - \gamma_0^{\nu}),$$

and

$$\Delta_X^{\nu} = (E[X|D=1] - E[X|D=0])^{\intercal}\gamma_0^{\nu}.$$

This particular decomposition is very easy to compute since it is similar to a standard Oaxaca decomposition. Firpo, Fortin and Lemieux (2007) point out, however, that there may be a bias in the decomposition because the linear specification used in the regression is only a local approximation that does not generally hold for larger changes in the covariates. A related point was made by Barsky et al. (2002) in the context of the Oaxaca decomposition for the mean. Barsky et al. point out that when the true conditional expectation is not linear, the decomposition based on a linear regression is biased. They suggest using a reweighting procedure instead, though this is not fully applicable here since we also want to estimate the contribution of each individual covariate.

Firpo, Fortin and Lemieux (2007) suggest a solution to this problem based on an hybrid approach that involves both reweighting and RIF regressions. They show that the following decomposition is valid:

$$\Delta_S^\nu = E\left[X|D=1\right]^\intercal \left(\gamma_1^\nu - \gamma_{01}^\nu\right), \tag{23}$$

and

$$\Delta_X^\nu = \left(E\left[X|D=1\right] - E\left[X|D=0\right]\right)^\intercal \gamma_0^\nu + \Delta_{X,e}^\nu, \tag{24}$$

where γ_{01}^ν are the coefficients from a RIF regression on the period 0 sample reweighted to have the same distribution of covariates as in period 1. The idea is that since a regression is the best linear approximation for a given distribution of X, this approximation may change when the distribution of X changes even if the wage structure remains the same. For example, if the true relationship between Y and a single X is convex, the linear regression coefficient will increase when we shift the distribution of X up, even if the true (convex) wage structure remains unchanged.

Back to our problem, this means that γ_1^ν and γ_0^ν may be different just because they are estimated for different distributions of X even if the wage structure remains unchanged over time. Since reweighting adjusts for this problem, we know for sure that $\gamma_1^\nu - \gamma_{01}^\nu$ reflects a true change in the wage structure. This is the reason why using $\gamma_1^\nu - \gamma_{01}^\nu$ for the decomposition yields a pure wage structure effect, while using $\gamma_1^\nu - \gamma_0^\nu$ instead would not necessarily do so.

Finally, in the case of the composition effect, the approximation (or specification) error term $\Delta_{X,e}^\nu$ is equal to the difference between the composition effect that would be consistently estimated using the reweighting approach, and the composition effect estimated using the regression-based linear approximation.

5.2 Results: RIF Regressions

Before showing the decomposition results, we first present some estimates from the RIF-regressions for the different wage quantiles, and for the variance of log wages and the Gini coefficient.[14] For wage percentiles, we compute $\text{IF}(y_i; q_\tau)$ for each observation using the sample estimate of q^τ, and the kernel density estimate of $f\left(q^\tau\right)$ using the Epanechnikov kernel and a bandwidth of 0.06. In addition to the reweighting factors discussed above,

[14] Firpo, Fortin, and Lemieux (2007) show the formula for the influence function in the case of the variance and the Gini coefficient.

we also use CPS sample weights throughout the empirical analysis. In practice, this means that we multiply the relevant reweighting factor with the CPS sample weight.

The list of covariates included in the regressions reflects the different explanations that have been suggested for the changes in the wage distribution over our sample period. Lemieux (2008) reviews possible explanations for the increased polarization in the labor market, including the technological-based explanation of Autor, Katz and Kearney. Furthermore, since it is well known that education wage differentials kept expanding after the late 1980s, the contribution of education to the wage structure effect is another leading explanation for inequality changes over this period.

Existing studies also indicate that composition effects played an important role over the 1988-2005 period. Lemieux (2006b) shows that all the growth in residual inequality over this period is due to composition effects linked to the fact that the workforce became older and more educated, two factors associated with more wage dispersion. Furthermore, Lemieux (2008) argues that de-unionization, another composition effect the way it is defined in this paper, contributed to the changes in the wage distribution over this period.

These various explanations can all be categorized in terms of the respective contributions of various sets of factors (occupational characteristics, unions, education, experience, etc.) to either wage structure or composition effects. This makes the decomposition method proposed in this paper ideally suited for estimating the contribution of each of these possible explanations to changes in the wage distribution. Applying our method to this issue fills an important gap in the literature, since no existing study has systematically attempted to estimate the contribution of each of the aforementioned factors to recent changes in the U.S. wage distribution.

As discussed earlier, our empirical analysis is based on data for men from the 1983-85 and 2000-02 ORG CPS. In light of the above discussion, the key set of covariates on which we focus are education (six education groups), potential experience (nine groups), union coverage, and the five measures of occupational task requirements introduced in Section 3. We also include controls for marital status and race in all the estimated models. The sample means for all these variables are provided in Table A1.

The RIF-regression coefficients for the 10^{th}, 50^{th}, and 90^{th} quantiles in 1988-90 and 2003-05, along with their (robust) standard errors are reported in Table 4. The RIF-regression coefficients for the variance and the Gini are reported in Table 5. We also report in Figure 3 and 4 the estimated coefficients from RIF-regressions for 19 different wage quantiles (from the 5^{th} to the 95^{th} quantile) equally spread over the whole wage distribution. This enables us to see whether different factors have different impacts at different points of the wage distribution. Using this flexible approach, as opposed to summary measures of inequality like the Gini coefficient or the variance of log wages, is important since wage dispersion changes very differently at different points of the distribution during this period.

Both Table 4 and the first panel of Figure 3 show that the effect of the union status across the different quantiles is highly non-monotonic. In both 1983-85 and 2000-2002, the effect first increases up to around the median, and then declines. The union effect even turns negative for the 90th and 95th quantiles. Overall, unions tend to reduce wage inequality, since the wage effect tends to be larger for lower than higher quantiles of the wage distribution. As shown by the RIF-regressions for the more global measures of inequality –the variance of log wages and the Gini coefficient– displayed in Table 5, the effect of unions on these measures is negative, although the magnitude of that effect has decreased over time. This is consistent with the well-known result (e.g. Freeman, 1980) that unions tend to reduce the variance of log wages for men.

More importantly, the results also indicate that unions *increase* inequality in the lower end of the distribution, but *decrease* inequality even more in the higher end of the distribution. For example, the estimates in Table 4 for 1983-85 imply that a 10 percent increase in the unionization rate would increase the 50-10 gap by 0.020, but decrease the 90-50 gap by 0.046.[15] As we will see later in the decomposition results, this means that the continuing decline in the rate of unionization can account for some of the "polarization" of the labor market (decrease in inequality at the low end, but increase in inequality at the top end).

The results for unions also illustrate an important feature of RIF regressions for quantiles, namely that they capture the effect of covariates on both the between- and within-group component of wage dispersion. The between-group effect dominates at the bottom end of the distribution, which explains why unions tend to increase inequality in that part of the distribution. The opposite happens, however, in the upper end of the wage distribution where the within-group effect dominates the between-group effect.

As in the case of unions, the RIF-regression estimates in Table 4 for other covariates also capture between- and within-group effects. Consider, for instance, the case of college education. Table 4 and Figure 3 show that the effect of college increases monotonically

[15] These numbers are obtained by multiplying the change in the unionization rate (0.1) by the difference between the effects at the 50th and 10th quantiles (0.406-0.208=0.198), and at the 90th and 50th quantiles (–0.055-0.406=-0.461).

as a function of percentiles. In other words, increasing the fraction of the workforce with a college degree has a larger impact on higher than lower quantiles. The reason why the effect is monotonic is that education increases both the level and the dispersion of wages (e.g. Lemieux, 2006a). As a result, both the within- and between-group effects go in the same direction of increasing inequality. Similarly, the effect of experience also tends to be monotonic as experience has a positive impact on both the level and the dispersion of wages.

Another clear pattern that emerges in Figures 3 and 4 is that, for most inequality enhancing covariates, i.e. those with a positively sloped curve, the inequality enhancing effect increases over time. In particular, the slopes for high levels of education (college graduates and post-graduates) become clearly steeper over time. This suggests that these covariates make a positive contribution to the wage structure effect.

The results for the five measures of occupational task are broadly consistent with the predictions of the technological change and offshoring literature. While the pattern of results in a given period are fairly non-linear and often hard to interpret, changes over time suggest an important role for these factors in changes in the distribution of wages. First consider the case of the information requirement of jobs. As in the case of unions, Figure 4 shows that this factor has an inverse U-shaped impact across the different percentiles of the wage distribution. More interestingly, the effect on wages declines in the lower middle of the distribution, but increase in the upper middle of the distribution. This may reflect the fact that this measure captures heterogenous tasks at different points of the distribution. For instance, we both use the O*NET scores on "processing information" and "analyzing data or information" to construct the information content measure. One possible explanation for the results is that workers processing information tend to earn lower wages and have experienced a decline in wages as their tasks are easier to execute with computers instead. By contrast, workers analyzing information may be earning higher wages, and have fared relatively well since the tasks they perform are complementary to computer technologies.

Changes over time in the wage effect of the second occupational task content measure, automation, indicate a large negative impact in the middle of the wage distribution, with a much smaller impact at the two ends of the distribution. This is consistent with Autor, Levy and Murnane (2003) who show that workers in the middle of the distribution are more likely to experience negative wage changes as the "routine" tasks they used to perform can now be executed by computer driven technologies instead.

Turning to the task measures linked to offshoring, Figure 4 shows that while jobs that involve no face-to-face contact tend to pay more than jobs that do involve face-to-face contact (positive RIF-regression estimated coefficients), this positive effect has declined substantially in the lower middle part of the wage distribution. This is consistent with the view that lower skill jobs that do not involve face-to-face contacts can be offshored, which has a negative impact on the wages of workers performing these kinds of tasks domestically. Similarly, the wage impact linked to jobs that do not involve on-site work has declined at the lower end of the wage distribution. Finally it is hard to discern any systematic pattern in the case of the decision-making measure. Jobs with no decision-making involved pay less at all points of the distribution, but the effect has not changed very much over time.

5.3 Decomposition Results

The results of the decomposition are presented in Figures 5-7. Tables 6 and 7 also summarize the results for the standard measure of top-end (90-50 gap) and low-end (50-10) wage inequality, as well as for the variance of log wages and the Gini coefficient. Note that the base group used in the RIF-regression models consists of non-union, white, and married men with a high school degree, 15 to 19 years of potential experience, and occupational task measures at one standard deviation below their sample averages. The covariates used in the RIF-regression models are those discussed above and listed in Table A1. A richer specification with additional interaction terms is used to estimate the logit models used compute the reweighting factor.[16]

The reweighting approach performs well in the sense that the reweighted means of the covariates for the 1983-85 sample are very close to those for 2000-02. Following Fortin, Lemieux, and Firpo (2010), we formally look at this by computing a reweighting error component of the decomposition linked to the fact that our logit model is not quite flexible enough to exactly reproduce the 2000-02 distribution of covariates.[17] Table 6 and Figure 7a show that this error is very small, which suggests that the logit reweighting model does closely replicate the distribution of covariates for 2000-02.

[16] The logit specification also includes a full set of interaction between experience and education, union status and education, union status and experience, and education and occupation task measures.

[17] To obtain the reweighting error, rewrite the wage structure effect as $\Delta_S^\nu = E[X|D=1]^\intercal \gamma_1^\nu - E[X|D=1]^\intercal \gamma_{01}^\nu$. In practice, we estimate this as $\widehat{\Delta}_S^\nu = \overline{X}_1^\intercal \widehat{\gamma}_1^\nu - \overline{X}_{01}^\intercal \widehat{\gamma}_{01}^\nu$, where $\overline{X}_{01}^\intercal$ and $\widehat{\gamma}_{01}^\nu$ are the means and regression coefficients in the reweighted sample, respectively. If the reweighting was replicating the means perfectly, we would have $\overline{X}_1^\intercal = \overline{X}_{01}^\intercal$. This means that the second term in equation $\widehat{\Delta}_S^\nu = \overline{X}_1^\intercal(\widehat{\gamma}_1^\nu - \widehat{\gamma}_{01}^\nu) + (\overline{X}_1^\intercal - \overline{X}_{01}^\intercal)\widehat{\gamma}_{01}^\nu$ would be equal to zero. This second term is the "reweighting error" that we report in Table 6 and Figure 7a. See Fortin, Lemieux, and Firpo (2010) for more detail.

As is well known (e.g. Oaxaca and Ransom, 1999), the detailed wage structure part of the decomposition (equation (23)) depends arbitrarily on the choice of the base group. This problem has mostly been discussed in the case of categorical variables, but it also applies in the case of continuous variables such as our task content measures.

To see this, consider $\widetilde{X}$, the "raw" variable, and X_B, the value of $\widetilde{X}$ chosen to be the base group. The base group is implicitly included in the regression model by using $X = \widetilde{X} - X_B$, the recentered version of $\widetilde{X}$, as a regressor. While X_B is often set to zero, we have in general that:

$$\begin{aligned}
\Delta_S^{\nu} &= E\left[X|D=1\right]^{\intercal}\left(\gamma_1^{\nu} - \gamma_{01}^{\nu}\right) \\
&= \left(E\left[\widetilde{X}|D=1\right] - X_B\right)^{\intercal}\left(\gamma_1^{\nu} - \gamma_{01}^{\nu}\right) \\
&\quad \left(E\left[\widetilde{X}|D=1\right] - X_B\right)^{\intercal}\gamma_1^{\nu} - \left(E\left[\widetilde{X}|D=1\right] - X_B\right)^{\intercal}\gamma_{01}^{\nu}.
\end{aligned}$$

The last line of the equation shows that the wage structure effect is the difference in the "treatment effect" of switching $\widetilde{X}$ from its base group value to its mean value under the wage structure parameters of period 1 (γ_1^{ν}) and 0 (γ_{01}^{ν}), respectively. Since the treatment effect obviously depends on the magnitude of the change in X ($E\left[\widetilde{X}|D=1\right] - X_B$), the wage structure effect also depends in an arbitrary way on the choice of X_B.

In the case of each task measure, we normalize the variable such that $E\left[\widetilde{X}|D=1\right] - X_B$ is equal to one standard deviation of the raw measure. The interpretation of the wage structure effect for each of these measures is thus the extent to which the wage impact of a one standard deviation increase in the measure has changed over time. But while the sign of the wage structure effect does not depend on the size of the increase in the X variable considered (one standard deviation), its magnitude does depend arbitrarily on this choice. The results presented below should, therefore, be interpreted with caution.

Figure 5a shows the overall change in (real log) wages at each percentile τ, (Δ_O^{τ}), and decomposes this overall change into a composition (Δ_X^{τ}) and wage structure (Δ_S^{τ}) effect using the reweighting procedure. Consistent with Autor, Katz and Kearney (2006), the overall change is U-shaped as wage dispersion increases in the top-end of the distribution, but declines in the lower end. This stands in sharp contrast with the situation that prevailed in the early 1980s when the corresponding curve was positively sloped as wage dispersion increased at all points of the distribution (Juhn, Murphy, and Pierce, 1993). Most summary measures of inequality such as the variance or the 90-10 gap nonetheless increase over the 1983-2002 period as wage gains in the top end of the distribution exceed those at the bottom end. In other words, though the curve for overall wage changes is

U-shaped, its slope is positive, on average, suggesting that inequality generally goes up.

Figure 5a also shows that, consistent with Lemieux (2006b), composition effects have contributed to a substantial increase in inequality. In fact, once composition effects are accounted for, the remaining wage structure effects follow a "purer" U-shape than unadjusted changes in wages. The lowest wage changes are now right in the middle of the distribution (from the 30^{th} to the 70^{th} percentile), while wage gains at the top and low end are quantitatively similar. Accordingly, Table 6 shows that all of the 0.051 change in the 90-10 gap is explained by the composition effects. By the same token, however, composition effects cannot account at all for the U-shaped nature of wage changes.

Figure 6 moves to the next step of the decomposition using RIF-regressions to attribute the contribution of each set of covariates to the composition effect. Figure 7 does the same for the wage structure effect. Figure 6a compares the "total" composition effect obtained by reweighting that was reported in Figure 5a, Δ_X^τ, to the composition effect explained using the RIF-regressions, $[\mathbb{E}[X|D=1]^\intercal - \mathbb{E}[X|D=0]]^\intercal \cdot \gamma_0^\tau$. The difference between the two curves is the specification (approximation) error $\Delta_{X,e}^\tau$. The error term is generally quite small and does not exhibit much of a systematic pattern. This means that the RIF-regression model does a good job at tracking down the composition effect estimated consistently using the reweighting procedure.

Figure 6b then divides the composition effect (explained by the RIF-regressions) into the contribution of five main sets of factors.[18] To simplify the discussion, let's focus on the impact of each factor in the lower and upper part of the distribution that is summarized in terms of the 50-10 and 90-50 gaps in Table 7. With the notable exception of unions, all factors have a larger impact on the 50-10 than on the 90-50 gap. The total contribution of all factors other than unionization is 0.041 and -0.007 for the 50-10 and 90-50 gaps, respectively. Composition effects linked to factors other than unions thus go the "wrong way" in the sense that they account for rising inequality at the bottom end while inequality is actually rising at the top end, a point noted earlier by Autor, Katz, and Kearney (2008).

In contrast, composition effects linked to unions (the impact of de-unionization) reduce inequality at the low end (effect of -0.016 on the 50-10) but increases inequality at the top end (effect of 0.038 on the 90-50). Note that, as in a Oaxaca decomposition, these effects on the 50-10 and the 90-50 gap can be obtained directly by multiplying the

[18] The effect of each set of factors is obtained by summing up the contribution of the relevant covariates. For example, the effect for "education" is the sum of the effect of each of the five education categories shown in Table A1. Showing the effect of each individual dummy separately would be cumbersome and harder to interpret.

8.2 percent decline in the unionization rate (Table A1) by the relevant union effects in 1983-85 shown in Table 4. The resulting effect of de-unionization accounts for 19 percent of the total change in the 50-10 gap, and 26 percent of the change in the 90-50 gap. The magnitude of these estimates is comparable to the relative contribution of de-unionization to the growth in inequality estimated for the 1980s (see Freeman, 1993, Card, 1992, and DiNardo, Fortin and Lemieux, 1996).

Figure 7a divides the wage structure effect, Δ_S^{τ}, into the part explained by the RIF-regression models, $\sum_{k=2}^{M} \mathbb{E}\left[X^k|D=1\right] \cdot \left[\gamma_{1,k}^{\nu} - \gamma_{01,k}^{\nu}\right]$, and the residual change $\gamma_{1,1}^{\nu} - \gamma_{01,1}^{\nu}$ (the change in the intercepts). The contribution of each set of factors is then shown in Figure 7b. As in the case of the composition effects, it is easier to discuss the results by focusing on the 90-50 and 50-10 gaps presented in Table 6 and 7. The results first show that -0.056 of the -0.136 change (decline) in the 50-10 gap due to wage structure effects remains unexplained (see the effect of the "constant" in Table 7). By contrast, covariates account for all of the 0.111 change in the 90-50 gap linked to the wage structure and even more, as the residual term (the constant) is now negative. Taken at face value, the results suggest a substantial decline in residual wage inequality at all points of the distribution. This finding may reflect, however, the peculiarities of the base group (see the discussion above).

Switching to the contribution of the different covariates, Table 7 shows that changes in the wage structure linked to education play a substantial role at the top end of the distribution, but do not have much impact at the lower end. These findings confirm Lemieux (2006a)'s conjecture that the large increase in the return to post-secondary education has contributed to a convexification of the wage distribution. Changes in the wage structure linked to experience go in the other direction, reflecting the fact that returns to experience have declined since the mid-1980s.

More importantly, the results show that changes in the wage structure linked to the technology and offshorability task measures have contributed to the U-shape change in the wage distribution over this period. Table 7 shows that both factors make a large and positive contribution to the increase in the 90-50 gap, and a sizable contribution to the decline in the 50-10 gap. Taken literally, the results suggest that these two factors can essentially account for all of the change in the 90-50 and 50-10 wage gap. This can also be seen in Figure 7b where the wage structure effects linked to technology and offshoring both follow a distinct U-shaped shape that closely mirrors the shape of the overall change in the wage distribution (Figure 5a). For the reasons mentioned above, however, the precise magnitude of the estimated effect is difficult to interpret because of the base group problem.

A number of interesting conclusions emerge, nonetheless, from the detailed wage decompositions. First, consistent with earlier studies, the composition effect linked to de-unionization accounts for about a fourth of the change in inequality both at the lower (50-10) and upper (90-50) end of the distribution. Second, the changing wage structure effects linked to unionization, education, and the occupational task measures of technology and offshoring all help explain the U-shaped feature of the changing wage distribution. Overall the results suggest that factors linked to occupation help explain some, but certainly not all, of the changes in the wage distribution observed between 1983-85 and 2000-02.

6 Conclusion

In this paper, we look at the contribution of occupations to changes in the distribution of wages. We first present a simple Roy-type skill pricing model, and use this as a motivation for estimating models for the change in within-occupation wage percentiles between 1983-85 and 2000-02. The findings from this first part of the empirical analysis suggest that changes in occupational wage profiles help explain the U-shape in changes in the wage distribution over this period. We also find that measures of technological change and offshorability at the occupation level help predict the changes in the occupational wage profiles.

We then explicitly quantify the contribution of these factors (technological change and offshoring) to changes in wage inequality relative to other explanations such as de-unionization and changes in the returns to education. We do so using a decomposition based on the influence function regression approach of by Firpo, Fortin, and Lemieux (2009). The results indicate that technological change and offshorability are two among a variety of other factors that can account for the observed changes in the distribution of wages.

More generally, our results suggest that even after controlling for standard skill measures, changes in both the level and dispersion of wages across occupations, as captured by our task measures, have played an important role in changes in the wage distribution. Our interpretation of this general finding is that returns to different dimensions of skill are different in different occupations, and have also changed differently over time. Like Acemoglu and Autor (2010), we conclude that it is essential to go beyond skills and formally introduce tasks and occupations in our standard models of the labor market to adequately understand why the wage distribution has changed so much over the last few decades.

REFERENCES

Autor, D.H., and D. Acemoglu (2010), "Skills, Tasks, and Technologies: Implications for Employment and Earnings" NBER Working Paper No. 16082 (forthcoming in O. Ashenfelter and D. Card, eds., *Handbook of Economics, vol. IV.)*

Autor, D.H., Katz, L.F. and M.S. Kearney (2008), "Trends in U.S. Wage Inequality: Revising the Revisionists," *Review of Economics and Statistics*, May

Autor, D.H., Katz, L.F. and M.S. Kearney (2006), "The Polarization of the U.S. Labor Market," *American Economic Review* 96, 189–194.

Autor, D.H., F. Levy, and R.J. Murnane (2003), "The Skill Content Of Recent Technological Change: An Empirical Exploration", *Quarterly Journal of Economics*, 118(4) 1279-1333.

Barsky, R., J. Bound, K. Charles, and J. Lupton (2002), "Accounting for the Black-White Wealth Gap: A Nonparametric Approach,"*Journal of the American Statistical Association*, 97(459), 663-673.

Blinder, A., (2007) "How Many U.S. Jobs Might Be Offshorable?," Center for Economic Policy Studies Working Discussion Paper no. 142, Princeton University.

Bound, J., and G. Johnson (1992), "Changes in the Structure of Wages in the 1980s: An Evaluation of Alternative Explanations," *American Economic Review* 82(3): 371-92

Card, D. (1992) "The Effects of Unions on the Distribution of Wages: Redistribution or Relabelling?" NBER Working Paper 4195, Cambridge: Mass.: National Bureau of Economic Research, 1992.

Crinò, R. (2009) "Service Offshoring and White-Collar Employment," UFAE and IAE Working Papers.

DiNardo, J., N.M. Fortin, and T. Lemieux, (1996), "Labor Market Institutions and the Distribution of Wages, 1973-1992: A Semiparametric Approach," *Econometrica*, 64, 1001-1044.

Firpo, S. Fortin, N.M., and T. Lemieux (2007) "Decomposing Distribution Using Recentered Influence Fucntion Regressions," Unpublished manuscript, PUC-Rio and UBC

Firpo, S. Fortin, N.M., and T. Lemieux (2009) "Unconditional Quantile Regressions," *Econometrica* 77(3), May, 953-973.

Firpo, S. Fortin, N.M., and T. Lemieux (2010) "Decomposition Methods in Economics," NBER Working Paper No. 16045 (forthcoming in O. Ashenfelter and D. Card, eds., *Handbook of Economics, vol. IV.)*

Feenstra, R, and G. Hanson (2003), "Global Production Sharing and Inequality: A Survey of Trade and Wages," in: Choi EK, Harrigan J (eds) *Handbook of International Trade*, Blackwell, Malden Oxford Victoria, 146-185.

Freeman, R.B. (1980), "Unionism and the Dispersion of Wages,' *Industrial and Labor Relations Review,* 34, 3-23.

Freeman, R.B.. (1993), "How Much has Deunionization Contributed to the Rise of Male Earnings Inequality?' In Sheldon Danziger and Peter Gottschalk, eds. *Uneven Tides: Rising Income Inequality in America.* New York: Russell Sage Foundation, 133-63.

Freeman, R.B.. (1995), "Are Your Wages Set in Beijing?" *Journal of Economic Perspectives*, 9(3), 15-32.

Gabaix, X., and A. Landier (2008) "Why Has CEO Pay Increased So Much?" *Quarterly Journal of Economics* 123(1), February, 49-100.

Gibbons, R., L.F. Katz, T. Lemieux, and D. Parent "Comparative Advantage, Learning, and Sectoral Wage Determination," *Journal of Labor Economics* 23(4), October, 681-724

Goos, Maarten and Alan Manning (2007) "Lousy and Lovely Jobs: The Rising Polarization of Work in Britain", *Review of Economics and Statistics*, 89(1), 118-133

Goos, Maarten, Alan Manning, and Anna Salomons (2009) "Recent Changes in the European Employment Structure: The Roles of Technological Change, Globalization and Institutions," unpublished manauscript.

Heckman, J., and J. Scheinkman (1987), "The Importance of Bundling in a Gorman-Lancaster Model of Earnings," *Review of Economic Studies*, 54(2), 243-55.

Jensen, J. Braford and Lori G. Kletzer, (2007) "Measuring the Task Content of Offshorable Services Jobs, Tradable Services and Job Loss," Peterson Institute for International Economics (shorter version forthcoming in *Labor in the New Economy*, Katharine Abraham, Mike Harper, and James Spletzer, eds., University of Chicago Press.)

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

Kaplan, S.N., and J. Rauh (2009) "Wall Street and Main Street: What Contributes to the Rise in the Highest Incomes?" *Review of Financial Studies*.

Lemieux, T., (2006a), "Post-secondary Education and Increasing Wage Inequality", *American Economic Review* 96(2), 195-199.

Lemieux, T., (2006b),"Increasing Residual Wage Inequality: Composition Effects, Noisy Data, or Rising Demand for Skill?", *American Economic Review* 96(3), 461-498.

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

Murphy, K.M., and F. Welch (1991) "The role of international trade in wage differentials," in M. Kosters (ed.) *Workers and Their Wages*, American Enterprise Institute Press, Washington DC, 39-69.

Oaxaca, R. and M.R. Ransom (1999), "Identification in Detailed Wage Decompositions," *Review of Economics and Statistics* 81(1), 154–157.

Welch, F. (1969) "Linear Synthesis of Skill Distribution," *The Journal of Human Resources*, 4(3), Summer, 311-327 .

Figure 1. Changes in Real Wages by Percentile, Men

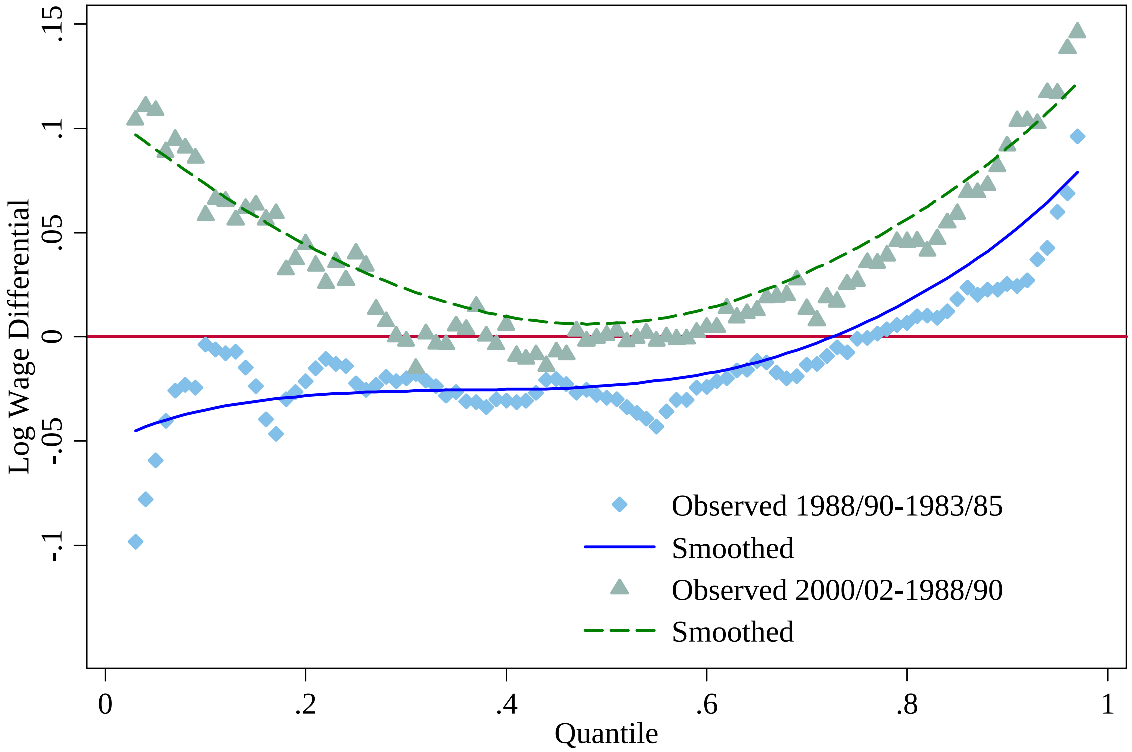


Figure 2a: Change in wage by 2-digit occupation

1983-85 to 2000-02 change for each decile

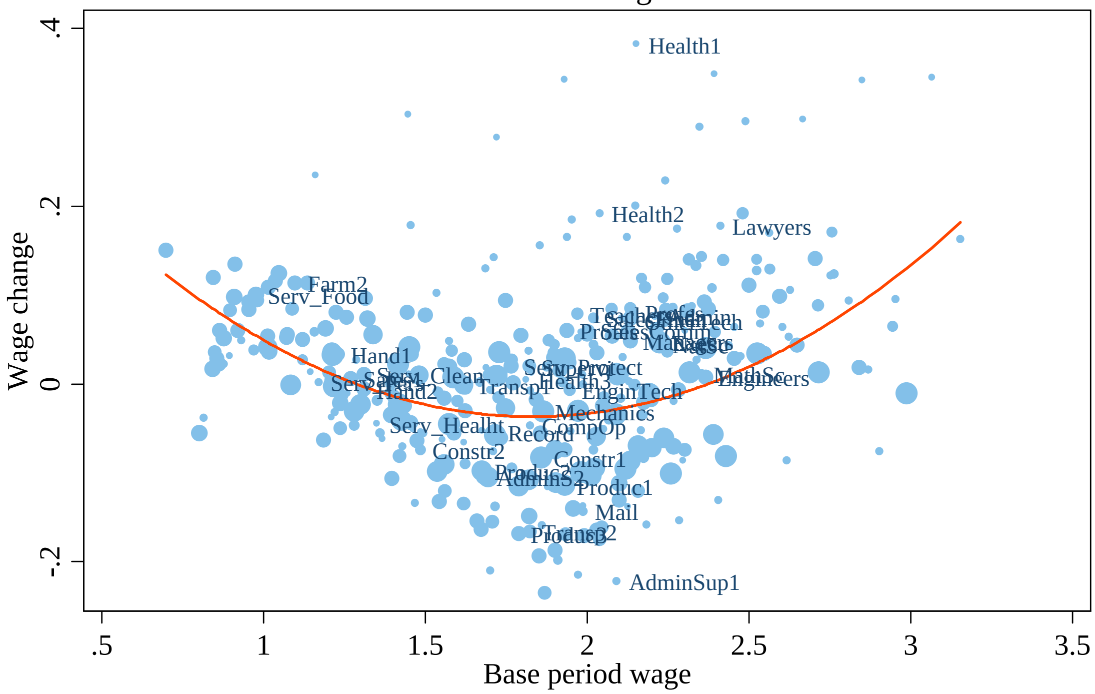


Figure 2b: Fitted change in wage by 2-digit occupation

1983-85 to 2000-02 change for each decile

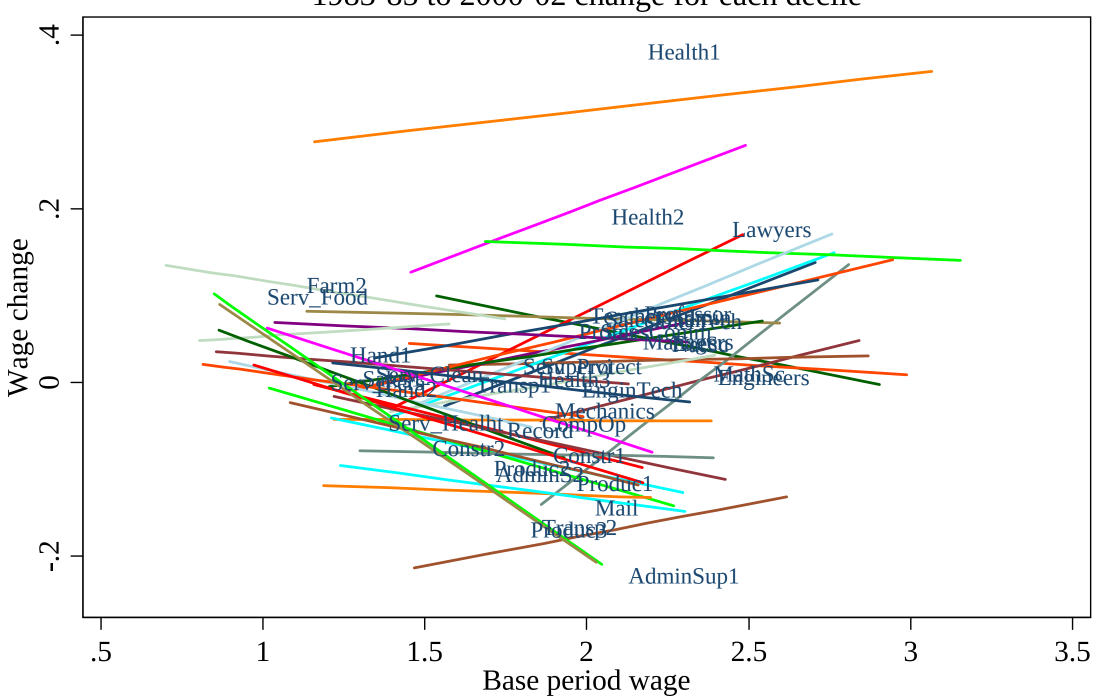


Figure 3. Unconditional Quantile Regressions Coefficients:
Selected Demographic Variables 1983/85-2000/02

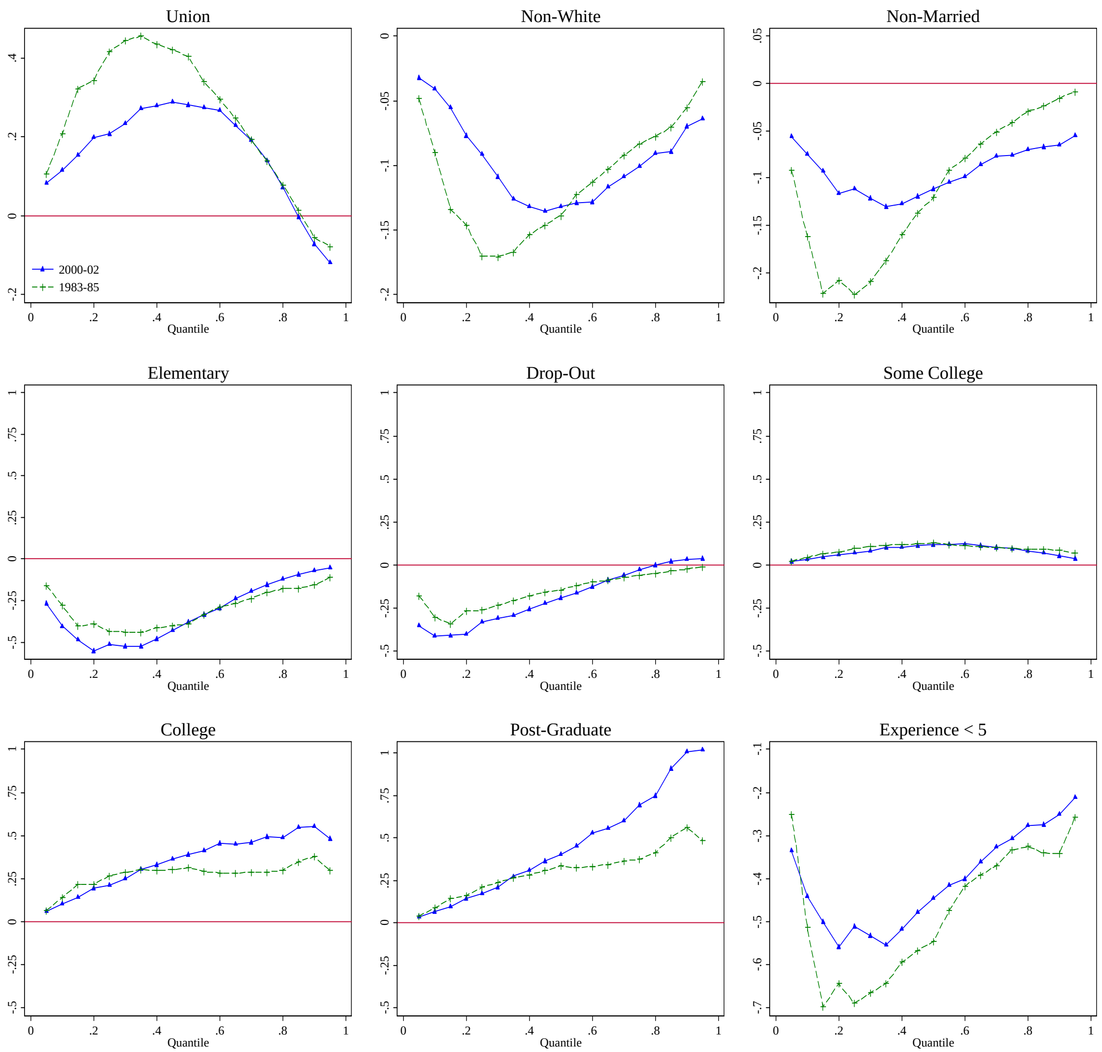


Figure 4. Unconditional Quantile Regressions Coefficients:
Occupational Tasks 1983/85-2000/02

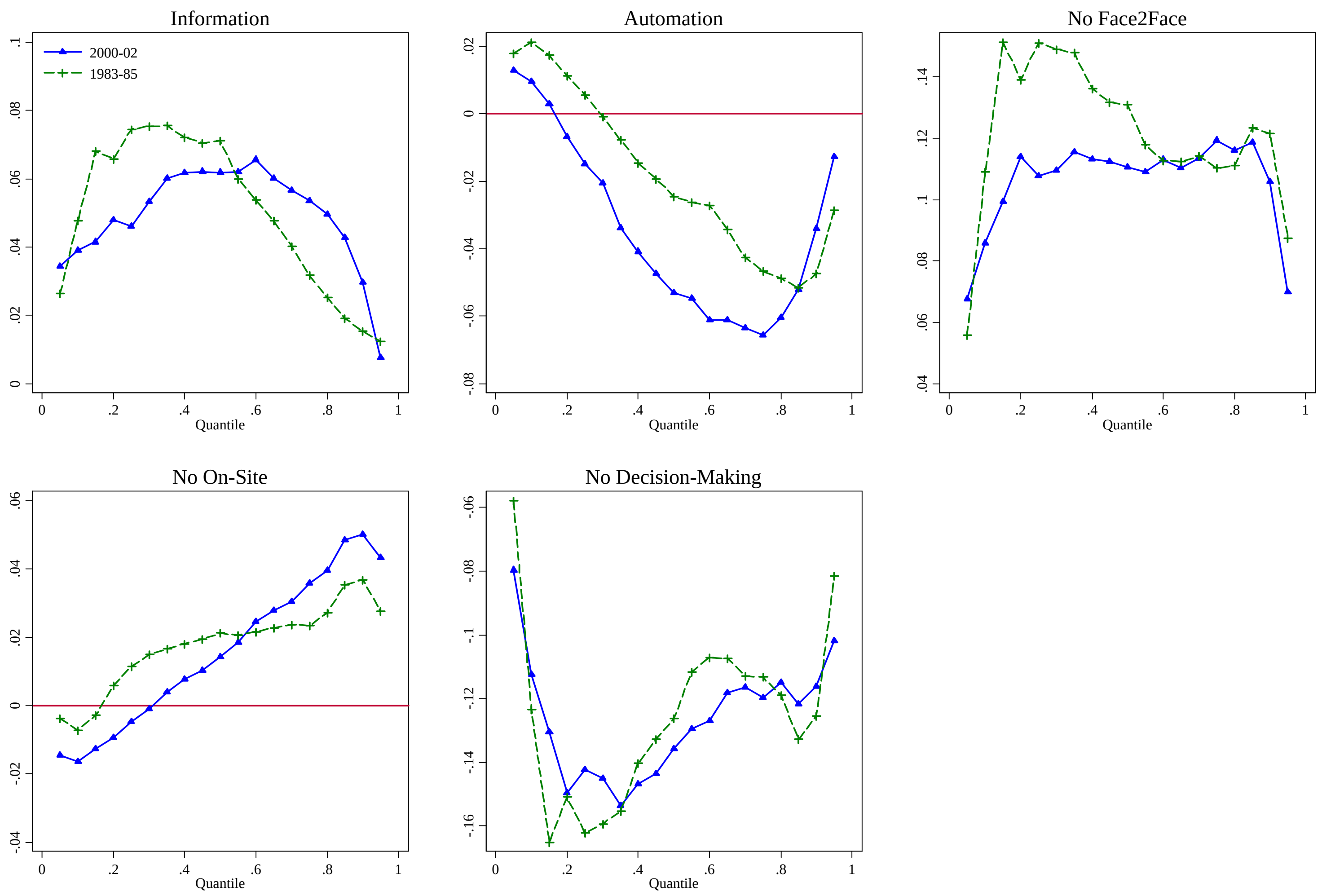
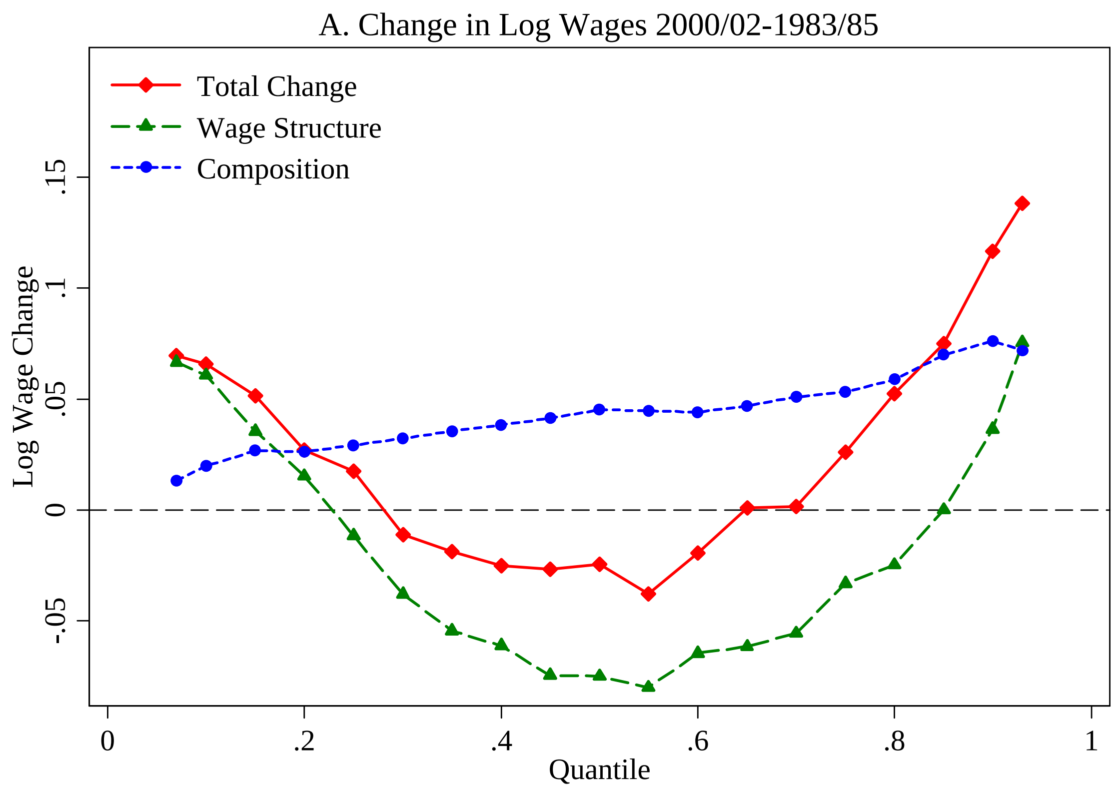


Figure 5. Decomposition of Total Change into Composition and Wage Structure Effects



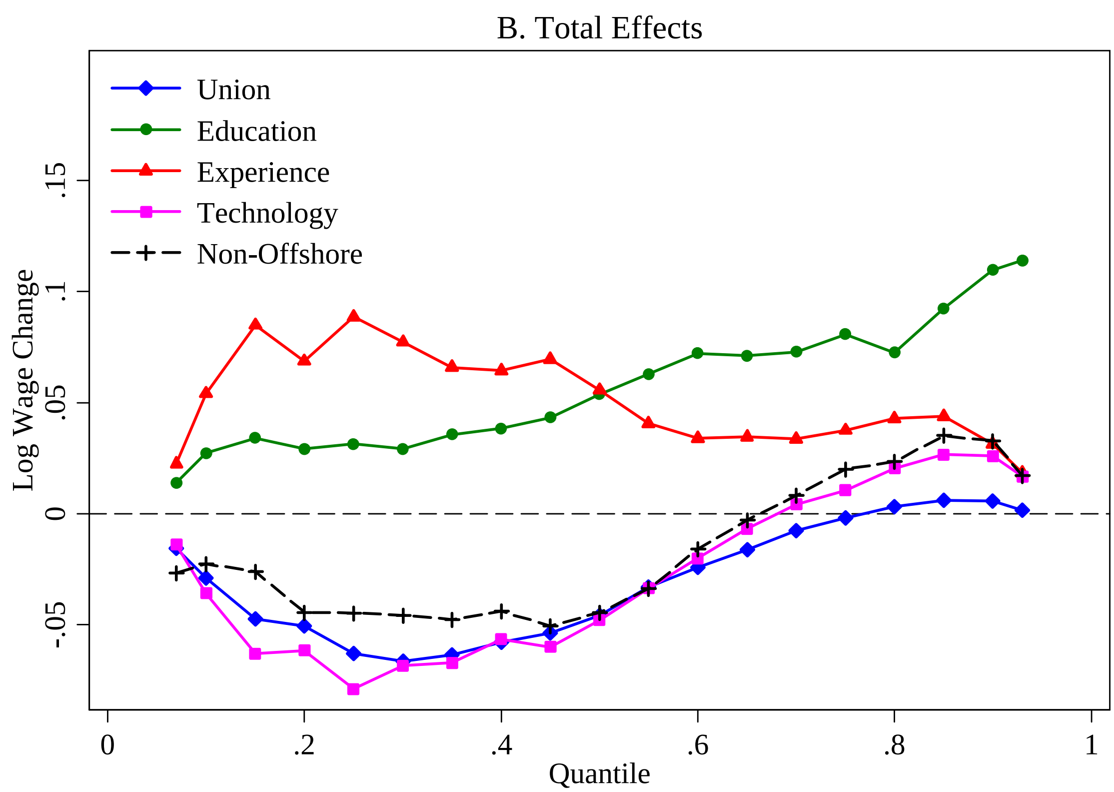
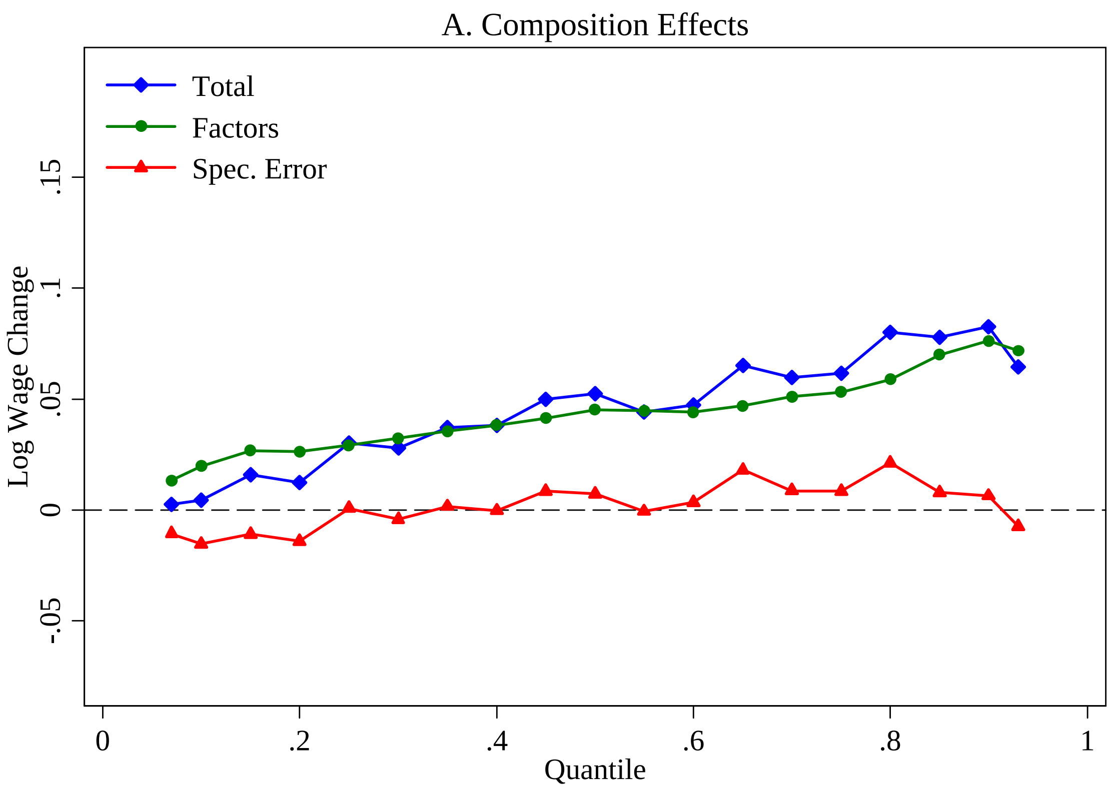


Figure 6. Decomposition of Composition Effects



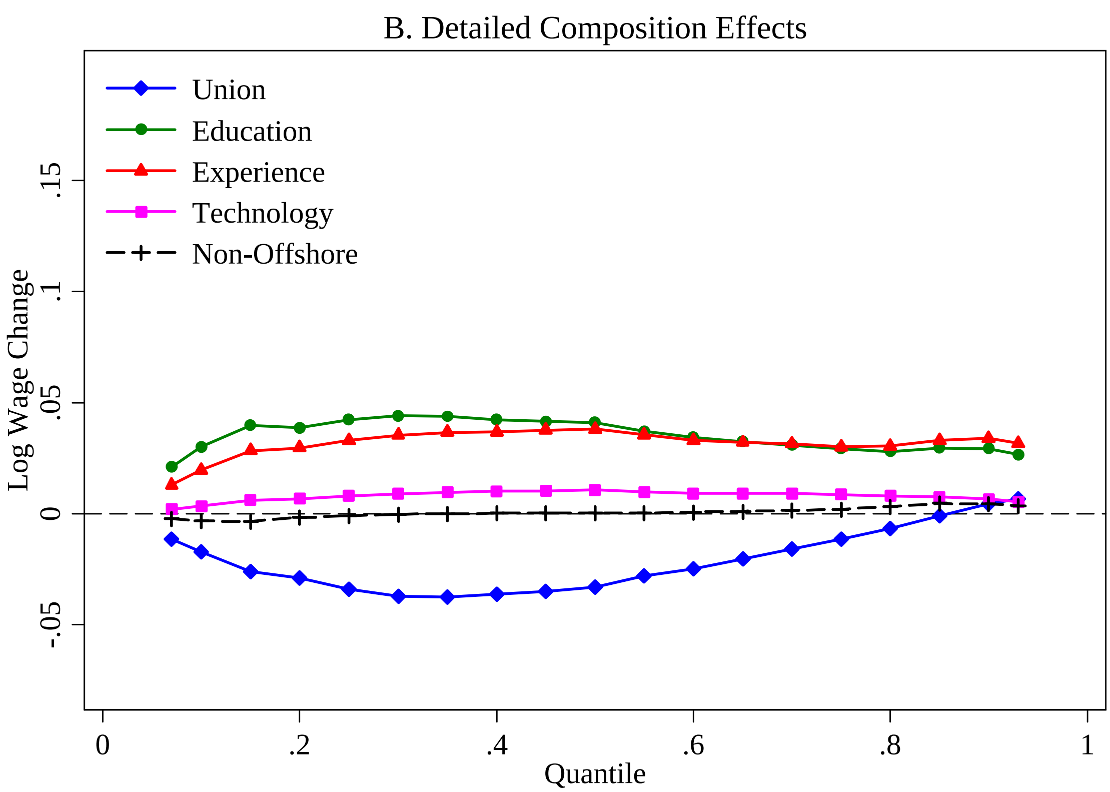
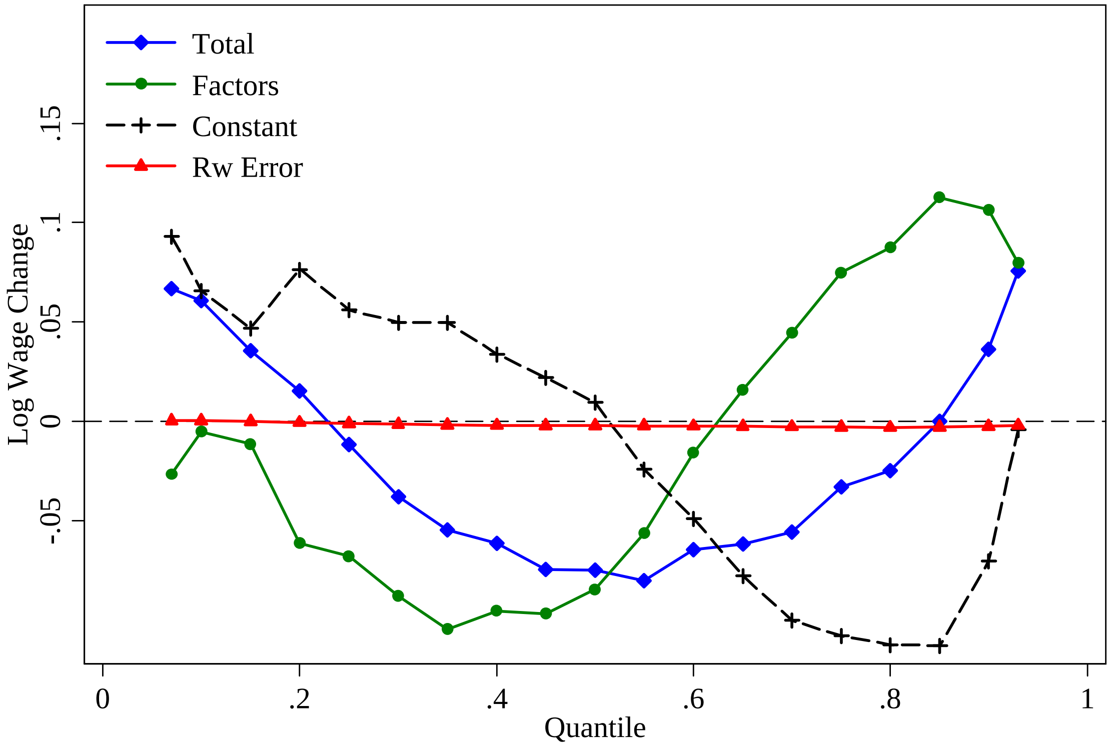


Figure 7. Decomposition of Wage Structure Effects

A. Wage Structure Effects



B. Detailed Wage Structure Effects

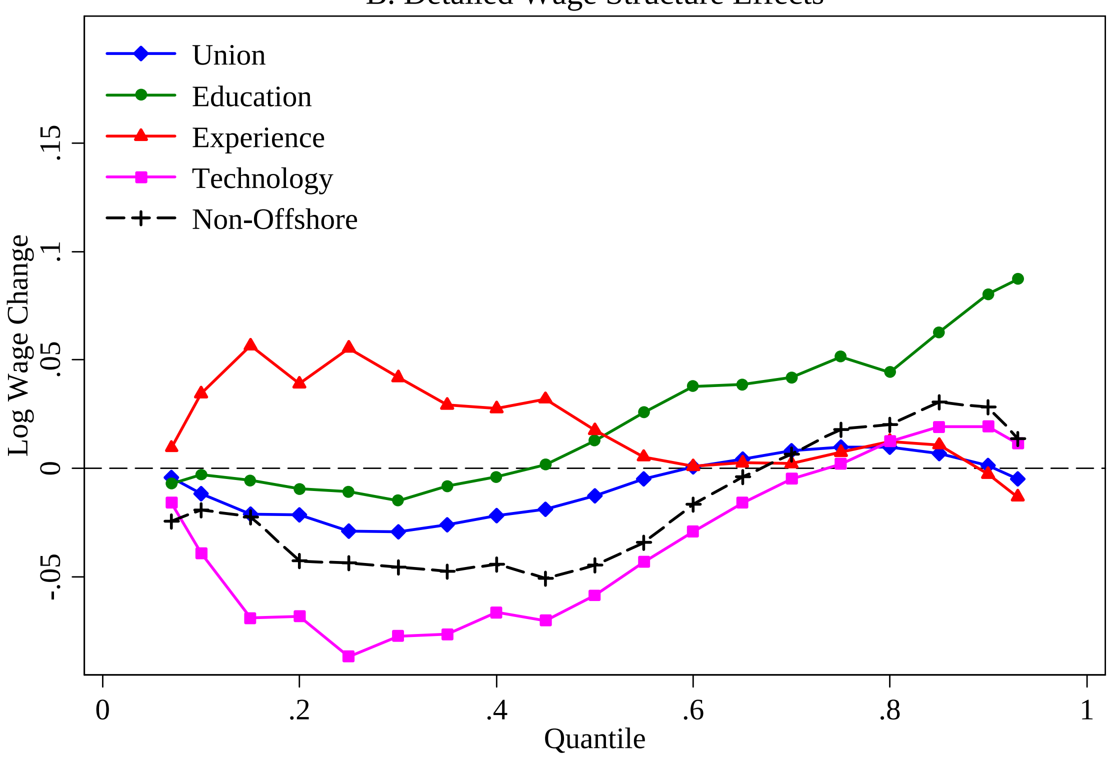


Table 1. O*NET 13.0 — Work Activities & Work Context

A. Characteristics linked to Technological Change/Offshorability[1]

Information Content

4.A.1.a.1	Getting Information (JK)
4.A.2.a.2	Processing Information (JK)
4.A.2.a.4	Analyzing Data or Information (JK)
4.A.3.b.1	Interacting With Computers (JK)
4.A.3.b.6	Documenting/Recording Information (JK)

Automation/Routinization

4.C.3.b.2	Degree of Automation
4.C.3.b.7	Importance of Repeating Same Tasks
4.C.3.b.8	Structured versus Unstructured Work (reverse)
4.C.3.d.3	Pace Determined by Speed of Equipment
4.C.2.d.1.i	Spend Time Making Repetitive Motions

B. Characteristics linked to Non-Offshorability

Face-to-Face

4.C.1.a.2.l	Face-to-Face Discussions
4.A.4.a.4	Establishing and Maintaining Interpersonal Relationships (JK,B)
4.A.4.a.5	Assisting and Caring for Others (JK,B)
4.A.4.a.8	Performing for or Working Directly with the Public (JK,B)
4.A.4.b.5	Coaching and Developing Others (B)

On-Site Job

4.A.1.b.2	Inspecting Equipment, Structures, or Material (JK)
4.A.3.a.2	Handling and Moving Objects
4.A.3.a.3	Controlling Machines and Processes
4.A.3.a.4	Operating Vehicles, Mechanized Devices, or Equipment
4.A.3.b.4	Repairing and Maintaining Mechanical Equipment (*0.5)
4.A.3.b.5	Repairing and Maintaining Electronic Equipment (*0.5)

Decision-Making

4.A.2.b.1	Making Decisions and Solving Problems (JK)
4.A.2.b.2	Thinking Creatively (JK)
4.A.2.b.4	Developing Objectives and Strategies
4.C.1.c.2	Responsibility for Outcomes and Results
4.C.3.a.2.b	Frequency of Decision Making

[1] Note: (JK) indicates a work activity used in Jensen and Kletzer (2007), (B) a work activity used or suggested in Blinder (2007).

Table 2: Regression fit of models for 1983-85 to 2000-02 changes in wages at each decile, by 2-digit occupation

	(1)	(2)	(3)	(4)	(5)	(6)
A. Models without controls for observables						
Adj. R-square	0.0023	0.6339	0.0060	0.6569	0.9310	0.8508
Curvature	0.125	0.065	0.114	0.054	-0.001	0.014
in residuals	(0.013)	(0.008)	(0.014)	(0.008)	(0.003)	(0.005)
B. Models with controls for observables						
Adj. R-square	0.0486	0.3915	0.0360	0.4420	0.8498	0.6738
Curvature	0.142	0.116	0.095	0.068	-0.001	0.044
in residuals	(0.016)	(0.012)	(0.017)	(0.012)	(0.006)	(0.009)
Occupation dummies		X		X	X	X
Decile dummies			X	X	X	
Base wage	X				X	X
Occ * base wage					X	X

Notes: Regression models estimated for each decile (10th, 20th,., 90th) of each 2-digit occupation. 360 observations used in all models (40 occupations, 9 observations per occupation). Models are weighted using the fraction of observations in the 2-digit occupation in the base period. Panel A shows the results when regressions are estimated without any controls for observabable. Panel B shows the results when the distribution of observables (age, education, race and marital status) in each occupation is reweighted to be the same as the overall distribution over all occupations.

Table 3: Estimated Effect of Task Requirements on Intercept and Slope of Wage Change Regressions by 2-digit Occupation

	Tasks included together						Included separately	
	Intercept			Slope			Intercept	Slope
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Information content	0.009	0.038	-0.007	0.016	-0.008	-0.003	0.050	0.042
	(0.013)	(0.014)	(0.013)	(0.014)	(0.015)	(0.022)	(0.013)	(0.016)
Automation /routine	-0.036	-0.052	-0.033	-0.026	-0.013	-0.022	-0.061	-0.056
	(0.014)	(0.013)	(0.012)	(0.015)	(0.014)	(0.021)	(0.011)	(0.013)
No on-site work	0.004	0.002	0.007	0.021	0.022	0.026	0.024	0.032
	(0.007)	(0.006)	(0.006)	(0.008)	(0.007)	(0.010)	(0.006)	(0.005)
No face-to-face	-0.048	-0.004	0.024	-0.002	-0.037	-0.064	-0.060	-0.071
	(0.018)	(0.019)	(0.018)	(0.019)	(0.022)	(0.031)	(0.013)	(0.013)
No decision making	0.037	0.001	-0.040	-0.034	-0.005	0.027	-0.053	-0.053
	(0.020)	(0.020)	(0.018)	(0.021)	(0.022)	(0.032)	(0.017)	(0.019)
Base wage	no	yes	yes	no	yes	yes	yes	yes
Reweighted	no	no	yes	no	no	yes	no	no
Adj. R-square	0.44	0.60	0.51	0.69	0.74	0.67	---	---

Notes: All models are estimated by running regressions of the 40 occupation-specific intercepts and slopes (estimated in specification (5) of Table 2) on the task measures. The models reported in all columns are weighted using the fraction of observations in each occupation in the base period. The intercepts and slopes used in column 3 and 6 are based on the regression models in Panel B of Table 2 where observables (age, education, race and marital status) are reweighted to be as in the overall distribution for all occupations. The models reported in all other columns rely on the estimates of Panel A that do not control for observables.

Table 4. Unconditional Quantile Regression Coefficients on Log Wages

Explanatory Variables	Years: Quantiles:	1983/85 10	50	90	2000/02 10	50	90
Union covered		0.208	0.406	-0.055	0.112	0.278	-0.073
		(0.003)	(0.004)	(0.004)	(0.003)	(0.005)	(0.006)
Non-white		-0.090	-0.140	-0.055	-0.040	-0.131	-0.071
		(0.006)	(0.004)	(0.004)	(0.006)	(0.004)	(0.006)
Non-Married		-0.162	-0.122	-0.015	-0.072	-0.111	-0.066
		(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.005)
Education (High School omitted)							
Primary		-0.278	-0.392	-0.156	-0.390	-0.378	-0.070
		(0.008)	(0.007)	(0.004)	0.012	(0.009)	(0.005)
Some HS		-0.301	-0.146	-0.023	-0.396	-0.188	0.034
		(0.007)	(0.004)	(0.004)	(0.01)	(0.005)	(0.004)
Some College		0.045	0.129	0.086	0.031	0.118	0.053
		(0.005)	(0.005)	(0.004)	(0.004)	(0.004)	(0.004)
College		0.142	0.316	0.375	0.102	0.386	0.561
		(0.004)	(0.006)	(0.007)	(0.004)	(0.005)	(0.011)
Post-grad		0.088	0.337	0.559	0.066	0.403	1.025
		(0.005)	(0.006)	(0.011)	(0.004)	(0.006)	(0.018)
Potential Experience (15< Experience < 20 omitted)							
Experience <5		-0.512	-0.549	-0.339	-0.427	-0.441	-0.254
		(0.009)	(0.007)	(0.007)	(0.008)	(0.007)	(0.009)
5< Experience < 10		-0.073	-0.318	-0.295	-0.080	-0.257	-0.265
		(0.005)	(0.007)	(0.008)	(0.005)	(0.007)	(0.01)
10< Experience < 15		-0.032	-0.157	-0.199	-0.029	-0.127	-0.136
		(0.004)	(0.005)	(0.008)	(0.005)	(0.006)	(0.01)
20< Experience < 25		-0.022	-0.054	-0.074	-0.012	-0.048	-0.017
		(0.005)	(0.006)	(0.008)	(0.005)	(0.005)	(0.011)
25< Experience < 30		-(0.005)	(0.028)	(0.048)	(0.)	(0.012)	(0.006)
		(0.005)	(0.006)	(0.01)	(0.004)	(0.005)	(0.01)
30< Experience < 35		0.002	0.034	0.055	0.001	0.015	0.021
		(0.005)	(0.006)	(0.009)	(0.005)	(0.006)	(0.01)
35< Experience < 40		0.017	0.031	0.056	0.005	0.003	0.037
		(0.005)	(0.008)	(0.009)	(0.005)	(0.007)	(0.012)
Experience > 40		0.055	0.011	-0.025	-0.002	-0.031	-0.029
		(0.006)	(0.007)	(0.008)	(0.008)	(0.008)	(0.012)
O*NET Measures							
Information Content		0.048	0.072	0.015	0.038	0.061	0.030
		(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)
Automation		0.021	-0.025	-0.047	0.009	-0.053	-0.035
		(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)
No Face-to-Face		0.109	0.132	0.121	0.083	0.110	0.107
		(0.003)	(0.002)	(0.003)	(0.002)	(0.002)	(0.004)
Non On-Site Job		-0.007	0.021	0.037	-0.016	0.014	0.051
		(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
No Decision-Making		-0.123	-0.127	-0.125	-0.109	-0.135	-0.118
		(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.004)
Number of obs.			274,625			252,397	

Note: Boostrapped standard errors (100 reps) are in parentheses.

Table 5. RIF Regression of Inequality Measures on Log Wages

	Variance:	0.2936	0.3380	Gini:	0.1750	0.1834
Explanatory Variables	Years:	1983/85	2000/02		1983/85	2000/02
Constant		0.2814	0.2359		0.1687	0.1538
		(0.0028)	(0.004)		(0.0009)	(0.001)
Union covered		-0.0974	-0.0747		-0.0568	-0.0350
		(0.0015)	(0.0023)		(0.0004)	(0.0006)
Non-white		0.0185	-0.0090		0.0171	0.0067
		(0.0025)	(0.0029)		(0.0008)	(0.0007)
Non-Married		0.0669	0.0169		0.0318	0.0135
		(0.0017)	(0.0024)		(0.0006)	(0.0006)
Education (High School omitted)						
Primary		0.0682	0.1459		0.0517	0.0744
		(0.0035)	(0.0046)		(0.0013)	(0.0018)
Some HS		0.0907	0.1576		0.0429	0.0646
		(0.0025)	(0.0033)		(0.0007)	(0.0011)
Some College		0.0125	0.0074		-0.0057	-0.0052
		(0.002)	(0.0021)		(0.0006)	(0.0007)
College		0.0609	0.1547		-0.0085	0.0104
		(0.003)	(0.0037)		(0.0008)	(0.0008)
Post-grad		0.1389	0.3306		0.0128	0.0491
		(0.0035)	(0.0053)		(0.0009)	(0.0013)
Potential Experience (15< Experience < 20 omitted)						
Experience <5		0.0791	0.0880		0.0734	0.0652
		(0.0032)	(0.0041)		(0.0011)	(0.001)
5< Experience < 10		-0.0633	-0.0569		0.0041	0.0042
		(0.0026)	(0.0039)		(0.0009)	(0.001)
10< Experience < 15		-0.0539	-0.0350		-0.0046	-0.0005
		(0.0023)	(0.004)		(0.0008)	(0.0008)
20< Experience < 25		-0.0189	-0.0036		-0.0011	0.0023
		(0.0027)	(0.004)		(0.0009)	(0.0009)
25< Experience < 30		0.0167	0.0025		0.0030	0.0000
		(0.0032)	(0.0038)		(0.001)	(0.0009)
30< Experience < 35		0.0162	0.0101		0.0022	0.0018
		(0.0036)	(0.0043)		(0.001)	(0.001)
35< Experience < 40		0.0148	0.0245		0.0015	0.0054
		(0.0034)	(0.005)		(0.0011)	(0.0011)
Experience > 40		-0.0238	0.0087		-0.0096	0.0045
		(0.0035)	(0.0056)		(0.001)	(0.0015)
O*NET Measures						
Information Content		-0.0147	-0.0076		-0.0097	-0.0065
		(0.0008)	(0.0013)		(0.0003)	(0.0003)
Automation		-0.0255	-0.0165		-0.0062	-0.0022
		(0.0009)	(0.0011)		(0.0003)	(0.0003)
No Face-to-Face		-0.0065	-0.0047		-0.0138	-0.0107
		(0.0012)	(0.0014)		(0.0004)	(0.0004)
Non On-Site Job		0.0106	0.0225		0.0017	0.0051
		(0.0005)	(0.0007)		(0.0002)	(0.0002)
No Decision-Making		0.0069	0.0070		0.0145	0.0142
		(0.0011)	(0.0018)		(0.0004)	(0.0005)

Note: Boostrapped standard errors (100 reps) are in parentheses.

Table 6. Aggregate Decomposition Results 1983/85-2000/02

Inequality Measure:	90-10	50-10	90-50	Variance	Gini
A: Specification 1					
Total Change	0.0511	-0.0900	0.1411	0.0443	0.0085
$\hat{\Delta}_O^{\nu} = \overline{\hat{RIF}(Y_1, \nu)} - \overline{\hat{RIF}(Y_0, \nu)}$	(0.0042)	(0.0035)	(0.0029)	(0.0015)	(0.0005)
Wage Structure	-0.0242	-0.1356	0.1113	0.0209	0.0055
$\hat{\Delta}_{S,p}^{\nu} = \overline{X}_1(\hat{\gamma}_1^{\nu} - \hat{\gamma}_{01}^{\nu})$	(0.0076)	(0.006)	(0.0064)	(0.0015)	(0.0004)
Composition	0.0563	0.0254	0.0310	0.0175	0.0012
$\hat{\Delta}_{X,p}^{\nu} = (\overline{X}_{01} - \overline{X}_0)\hat{\gamma}_0^{\nu}$	(0.0006)	(0.0004)	(0.0004)	(0.0005)	(0.0003)
Specification Error	0.0217	0.0227	-0.0009	0.0069	0.0018
$\hat{\Delta}_{X,e}^{\nu} = \overline{X}_{01}(\hat{\gamma}_{01}^{\nu} - \hat{\gamma}_0^{\nu})$	(0.0044)	(0.0026)	(0.0041)	(0.0006)	(0.0002)
Reweighting Error	-0.0027	-0.0025	-0.0002	-0.0010	-0.0001
$\hat{\Delta}_{S,e}^{\nu} = (\overline{X}_1 - \overline{X}_{01})\hat{\gamma}_{01}^{\nu}$	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)

Note: Boostrapped standard errors (100 reps) are in parentheses. Repetitions conducted over the entire procedure.

Table 7. Detailed Decomposition Results 1983/85-2000/02

Inequality Measure:	90-10	50-10	90-50	Variance	Gini
A: Detailed Wage Structure Effects					
Union	0.0131	-0.0008	0.0139	0.0045	0.0030
	(0.0002)	(0.0001)	(0.0001)	(0.0005)	(0.0001)
Education	0.0833	0.0157	0.0676	0.0316	0.0070
	(0.0015)	(0.0013)	(0.0015)	(0.0018)	(0.0006)
Experience	-0.0370	-0.0168	-0.0201	-0.0107	-0.0055
	(0.0028)	(0.0016)	(0.0017)	(0.0043)	(0.0012)
Technology	0.0586	-0.0194	0.0779	0.0233	0.0107
	(0.004)	(0.0032)	(0.0036)	(0.0026)	(0.0007)
Offshorability	0.0476	-0.0255	0.0731	0.0243	0.0090
	(0.0054)	(0.0052)	(0.0059)	(0.003)	(0.0007)
Other	-0.0539	-0.0325	-0.0214	-0.0213	-0.0074
	(0.0003)	(0.0002)	(0.0003)	(0.0016)	(0.0005)
Constant	-0.1359	-0.0563	-0.0796	-0.0309	-0.0116
	(0.0093)	(0.008)	(0.009)	(0.0064)	(0.0017)
Total	-0.0242	-0.1356	0.1113	0.0209	0.0055
	(0.0076)	(0.006)	(0.0064)	(0.0015)	(0.0004)
B: Detailed Composition Effects:					
Union	0.0217	-0.0159	0.0375	0.0080	0.0047
	(0.0005)	(0.0005)	(0.0007)	(0.0002)	(0.0001)
Education	-0.0008	0.0109	-0.0117	-0.0019	-0.0041
	(0.0011)	(0.0008)	(0.0008)	(0.0003)	(0.0001)
Experience	0.0142	0.0184	-0.0042	0.0033	-0.0021
	(0.001)	(0.0007)	(0.0009)	(0.0003)	(0.0001)
Technology	0.0032	0.0074	-0.0042	0.0009	-0.0005
	(0.0004)	(0.0004)	(0.0004)	(0.0003)	(0.0001)
Offshorability	0.0077	0.0035	0.0042	0.0023	0.0007
	(0.0009)	(0.0003)	(0.0004)	(0.0002)	(0.0003)
Other	0.0103	0.0010	0.0093	0.0048	0.0067
	(0.0009)	(0.0005)	(0.0004)	(0.0002)	(0.0001)
Total	0.0563	0.0254	0.0310	0.0175	0.0012
	(0.0006)	(0.0004)	(0.0004)	(0.0005)	(0.0003)

Note: Boostrapped standard errors (100 reps) are in parentheses. Repetitions conducted over the entire procedure.

Table A1. Descriptive Statistics

	1983-85		2003-05		Difference in Means
	Means	Standard Deviation	Means	Standard Deviation	
Log wages	1.767	0.542	1.797	0.581	0.029
Union covered	0.236	0.425	0.154	0.361	-0.082
Non-white	0.122	0.327	0.157	0.364	0.035
Non-Married	0.361	0.480	0.422	0.494	0.061
Education					
Primary	0.067	0.250	0.040	0.197	-0.027
Some HS	0.127	0.333	0.086	0.280	-0.041
High School					
Some College	0.195	0.396	0.272	0.445	0.078
College	0.130	0.336	0.184	0.388	0.054
Post-grad	0.097	0.296	0.091	0.287	-0.007
Age	35.766		38.249		2.483
O*NET Measures					
Information Content	1.419	1.419	1.518	1.446	0.100
Automation	1.990	1.990	0.961	1.012	-1.029
No Face-to-Face	0.911	0.911	0.806	0.919	-0.105
Non On-Site Job	1.050	1.050	2.162	1.977	1.112
No Decision-Making	1.045	1.045	0.968	1.061	-0.082

Appendix Table B1.

Examples of Work Activities for O*Net Occupation for (11-2022) Sales Managers and (15-1021) Computer Programmers and (15-1032) Computer Software Engineers (Systems)

11-2022.00	4.A.2.b.2	Thinking Creatively	IM	3.95
11-2022.00	4.A.2.b.2	Thinking Creatively	LV	4.25
11-2022.00	4.A.3.b.1	Interacting With Computers	IM	3.80
11-2022.00	4.A.3.b.1	Interacting With Computers	LV	3.60
11-2022.00	4.A.4.a.5	Assisting and Caring for Others	IM	2.25
11-2022.00	4.A.4.a.5	Assisting and Caring for Others	LV	2.30
15-1021.00	4.A.2.b.2	Thinking Creatively	IM	3.11
15-1021.00	4.A.2.b.2	Thinking Creatively	LV	4.01
15-1021.00	4.A.3.b.1	Interacting With Computers	IM	4.99
15-1021.00	4.A.3.b.1	Interacting With Computers	LV	5.39
15-1021.00	4.A.4.a.5	Assisting and Caring for Others	IM	2.86
15-1021.00	4.A.4.a.5	Assisting and Caring for Others	LV	3.21
15-1032.00	4.A.2.b.2	Thinking Creatively	IM	3.95
15-1032.00	4.A.2.b.2	Thinking Creatively	LV	5.34
15-1032.00	4.A.3.b.1	Interacting With Computers	IM	5.00
15-1032.00	4.A.3.b.1	Interacting With Computers	LV	6.11
15-1032.00	4.A.4.b.5	Coaching and Developing Others	IM	2.62
15-1032.00	4.A.4.b.5	Coaching and Developing Others	LV	2.71

Examples of Work Context for O*Net Occupation for (11-2022) Sales Managers and (15-1021) Computer Programmers and (15-1032) Computer Software Engineers (Systems)

4.C.3.a.2.b	Frequency of Decision Making	CXP	1 Never
4.C.3.a.2.b	Frequency of Decision Making	CXP	2 Once a year or more but not every month
4.C.3.a.2.b	Frequency of Decision Making	CXP	3 Once a month or more but not every week
4.C.3.a.2.b	Frequency of Decision Making	CXP	4 Once a week or more but not every day
4.C.3.a.2.b	Frequency of Decision Making	CXP	5 Every day

11-2022.00	4.C.3.a.2.b	Frequency of Decision Making	CX	n/a	3.62
11-2022.00	4.C.3.a.2.b	Frequency of Decision Making	CXP	1	0.00
11-2022.00	4.C.3.a.2.b	Frequency of Decision Making	CXP	2	19.05
11-2022.00	4.C.3.a.2.b	Frequency of Decision Making	CXP	3	33.3
11-2022.00	4.C.3.a.2.b	Frequency of Decision Making	CXP	4	14.29
11-2022.00	4.C.3.a.2.b	Frequency of Decision Making	CXP	5	33.33
15-1021.00	4.C.3.a.2.b	Frequency of Decision Making	CX	n/a	3.14
15-1021.00	4.C.3.a.2.b	Frequency of Decision Making	CXP	1	15.03
15-1021.00	4.C.3.a.2.b	Frequency of Decision Making	CXP	2	8.49
15-1021.00	4.C.3.a.2.b	Frequency of Decision Making	CXP	3	34.93
15-1021.00	4.C.3.a.2.b	Frequency of Decision Making	CXP	4	30.46
15-1021.00	4.C.3.a.2.b	Frequency of Decision Making	CXP	5	11.09

15-1032.00	4.C.3.a.2.b	Frequency of Decision Making	CX	n/a	3.46
15-1032.00	4.C.3.a.2.b	Frequency of Decision Making	CXP	1	0.00
15-1032.00	4.C.3.a.2.b	Frequency of Decision Making	CXP	2	30.49
15-1032.00	4.C.3.a.2.b	Frequency of Decision Making	CXP	3	4.63
15-1032.00	4.C.3.a.2.b	Frequency of Decision Making	CXP	4	53.35
15-1032.00	4.C.3.a.2.b	Frequency of Decision Making	CXP	5	11.53

4.C.3.b.7	Importance of Repeating Same Tasks	CXP	1	Not important at all
4.C.3.b.7	Importance of Repeating Same Tasks	CXP	2	Fairly important
4.C.3.b.7	Importance of Repeating Same Tasks	CXP	3	Important
4.C.3.b.7	Importance of Repeating Same Tasks	CXP	4	Very important
4.C.3.b.7	Importance of Repeating Same Tasks	CXP	5	Extremely important

11-2022.00	4.C.3.b.7	Importance of Repeating Same Tasks	CX	n/a	2.67
11-2022.00	4.C.3.b.7	Importance of Repeating Same Tasks	CXP	1	19.05
11-2022.00	4.C.3.b.7	Importance of Repeating Same Tasks	CXP	2	28.57
11-2022.00	4.C.3.b.7	Importance of Repeating Same Tasks	CXP	3	28.57
11-2022.00	4.C.3.b.7	Importance of Repeating Same Tasks	CXP	4	14.29
11-2022.00	4.C.3.b.7	Importance of Repeating Same Tasks	CXP	5	9.52

15-1021.00	4.C.3.b.7	Importance of Repeating Same Tasks	CX	n/a	2.74
15-1021.00	4.C.3.b.7	Importance of Repeating Same Tasks	CXP	1	16.29
15-1021.00	4.C.3.b.7	Importance of Repeating Same Tasks	CXP	2	28.00
15-1021.00	4.C.3.b.7	Importance of Repeating Same Tasks	CXP	3	31.59
15-1021.00	4.C.3.b.7	Importance of Repeating Same Tasks	CXP	4	13.82
15-1021.00	4.C.3.b.7	Importance of Repeating Same Tasks	CXP	5	10.31

15-1032.00	4.C.3.b.7	Importance of Repeating Same Tasks	CX	n/a	
15-1032.00	4.C.3.b.7	Importance of Repeating Same Tasks	CXP	1	13.04
15-1032.00	4.C.3.b.7	Importance of Repeating Same Tasks	CXP	2	0.09
15-1032.00	4.C.3.b.7	Importance of Repeating Same Tasks	CXP	3	7.56
15-1032.00	4.C.3.b.7	Importance of Repeating Same Tasks	CXP	4	20.12
15-1032.00	4.C.3.b.7	Importance of Repeating Same Tasks	CXP	5	59.20

4.C.3.b.8	Structured versus Unstructured Work	CXP	1	No freedom
4.C.3.b.8	Structured versus Unstructured Work	CXP	2	Very little eedom
4.C.3.b.8	Structured versus Unstructured Work	CXP	3	Limited freedom
4.C.3.b.8	Structured versus Unstructured Work	CXP	4	Some freedom
4.C.3.b.8	Structured versus Unstructured Work	CXP	5	A lot of freedom

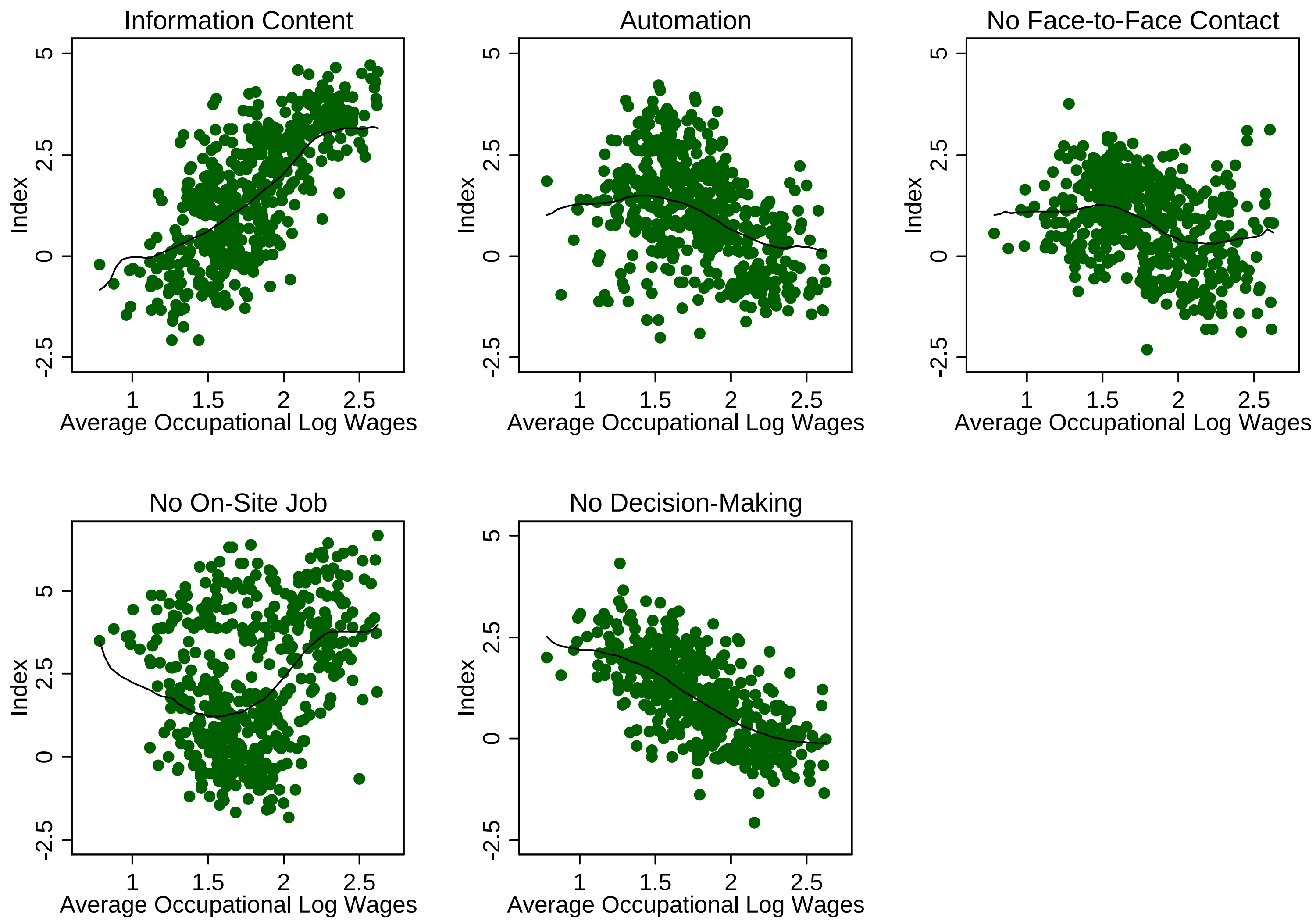
11-2022.00	4.C.3.b.8	Structured versus Unstructured Work	CX	n/a	4.33
11-2022.00	4.C.3.b.8	Structured versus Unstructured Work	CXP	1	0.00
11-2022.00	4.C.3.b.8	Structured versus Unstructured Work	CXP	2	4.76
11-2022.00	4.C.3.b.8	Structured versus Unstructured Work	CXP	3	0.00
11-2022.00	4.C.3.b.8	Structured versus Unstructured Work	CXP	4	52.38
11-2022.00	4.C.3.b.8	Structured versus Unstructured Work	CXP	5	42.86

15-1021.00	4.C.3.b.8	Structured versus Unstructured Work	CX	n/a	3.99
15-1021.00	4.C.3.b.8	Structured versus Unstructured Work	CXP	1	0.80
15-1021.00	4.C.3.b.8	Structured versus Unstructured Work	CXP	2	1.12
15-1021.00	4.C.3.b.8	Structured versus Unstructured Work	CXP	3	1.11
15-1021.00	4.C.3.b.8	Structured versus Unstructured Work	CXP	4	91.87
15-1021.00	4.C.3.b.8	Structured versus Unstructured Work	CXP	5	5.09

15-1032.00	4.C.3.b.8	Structured versus Unstructured Work	CX	n/a	4.57

15-1032.00	4.C.3.b.8	Structured versus Unstructured Work	CXP	1	1.52
15-1032.00	4.C.3.b.8	Structured versus Unstructured Work	CXP	2	0.28
15-1032.00	4.C.3.b.8	Structured versus Unstructured Work	CXP	3	0.21
15-1032.00	4.C.3.b.8	Structured versus Unstructured Work	CXP	4	35.29
15-1032.00	4.C.3.b.8	Structured versus Unstructured Work	CXP	5	62.70

Appendix Figure 1. Average Occupational Wages in 2000-02 and Task Category Indexes



Appendix Figure 2. Average Occupational Wages in 2000-02 and Autor et al. (2003) Indexes