



IZA DP No. 5542

## Occupational Tasks and Changes in the Wage Structure

Sergio Firpo  
Nicole M. Fortin  
Thomas Lemieux

February 2011

# Occupational Tasks and Changes in the Wage Structure

**Sergio Firpo**

*São Paulo School of Economics, FGV  
and IZA*

**Nicole M. Fortin**

*University of British Columbia  
and IZA*

**Thomas Lemieux**

*University of British Columbia*

Discussion Paper No. 5542  
February 2011

IZA

P.O. Box 7240  
53072 Bonn  
Germany

Phone: +49-228-3894-0  
Fax: +49-228-3894-180  
E-mail: [iza@iza.org](mailto:iza@iza.org)

Any opinions expressed here are those of the author(s) and not those of IZA. Research published in this series may include views on policy, but the institute itself takes no institutional policy positions.

The Institute for the Study of Labor (IZA) in Bonn is a local and virtual international research center and a place of communication between science, politics and business. IZA is an independent nonprofit organization supported by Deutsche Post Foundation. The center is associated with the University of Bonn and offers a stimulating research environment through its international network, workshops and conferences, data service, project support, research visits and doctoral program. IZA engages in (i) original and internationally competitive research in all fields of labor economics, (ii) development of policy concepts, and (iii) dissemination of research results and concepts to the interested public.

IZA Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

## ABSTRACT

### Occupational Tasks and Changes in the Wage Structure<sup>\*</sup>

This paper argues that changes in the returns to occupational tasks have contributed to changes in the wage distribution over the last three decades. Using Current Population Survey (CPS) data, we first show that the 1990s polarization of wages is explained by changes in wage setting between and within occupations, which are well captured by tasks measures linked to technological change and offshorability. Using a decomposition based on Firpo, Fortin, and Lemieux (2009), we find that technological change and deunionization played a central role in the 1980s and 1990s, while offshorability became an important factor from the 1990s onwards.

JEL Classification: J3, J5

Keywords: wage inequality, polarization, occupational tasks, offshoring, RIF-regressions

Corresponding author:

Sergio Firpo  
Escola de Economia de São Paulo – FGV-SP  
R. Itapeva, 474/1215  
São Paulo, SP 01332-000  
Brazil  
E-mail: [sergio.firpo@fgv.br](mailto:sergio.firpo@fgv.br)

---

<sup>\*</sup> We are indebted to David Autor, David Card, David Dorn, Martin Goos, Caroline Hoxby, Luigi Pistaferri, Anna Salomon, Alexandra Spitz-Oener, Frank Vella, and seminar participants at UC Berkeley, Cornell, Georgetown, UCL, LSE, HEC-Montreal, Ottawa, Toronto, Calgary, Hawaii, Paris II, PSE, Columbia, Stanford, Victoria, Collegio Carlo Alberto, UC Davis, the 2010 NBER Summer Institute, the T.A.S.K.S. International Workshop in Nuremberg, the 2010 EALE/SOLE Conference, and the ACAM Conference in Tokyo for useful comments on this and earlier versions of the manuscript. Financial support from CNPq-Brazil for Firpo, from CIFAR and SSHRC for Fortin, and from SSHRC for Lemieux was greatly appreciated. The usual disclaimer applies.

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).<sup>1</sup> Until recently, little attention had been paid to the potential role of occupations in changes in wage inequality. This situation has changed over the last five years for several 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, computer and information technologies have depressed the return to “routine” tasks that can now be executed by computer technologies. 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. Technological change could thus explain why wages in the middle of the distribution fell more than wages at the bottom and top end of the distribution.<sup>2</sup>

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.<sup>3</sup> 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 of technological change, that we discuss below, is that changes in the wage structure within and between occupations should be systematically related to 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. Early explanations for the role of international trade in changes in inequality have

<sup>1</sup>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).

<sup>2</sup>Acemoglu and Autor (2011) develop a formal model to show how this could happen in a model with three skill levels (high, middle, and low).

<sup>3</sup>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. Alternatively, Goos, Manning, and Salomon (2009) explore the role of offshoring in the polarization of employment in Europe using counts of news reports about offshoring of European jobs from the European Restructuring Monitor.

focused on the role of trade in final products that are defined at the industry level. It was later argued (Feenstra and Hanson, 2003) that trade in intermediate inputs was a more promising explanation for changes in wage inequality than trade in final goods and services. More recently, Grossman and Rossi-Hansberg (2008) have proposed a model where offshoring figures as a source of task supply among other competing sources. 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. This suggests assessing the role of offshoring in changes in the wage structure by contrasting the evolution of wages and employment in occupations that are potentially offshorable (e.g. Blinder (2007) and Jensen and Kletzer (2010)) relative to those that are not offshorable.

Although occupations now feature prominently as a possible channel 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 occupation-based explanations. 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 (2008) 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. Acemoglu and Autor (2011) 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 can precisely be accounted for by occupation-based explanations.

The goal of this paper is to fill this gap by systematically investigating the contribution of occupations to changes in the distribution of U.S. male wages.<sup>4</sup> 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

<sup>4</sup>We focus on men for which the phenomenon of polarization is thought to have the more dire effects (Autor, 2010).

offshorability. Using Current Population Survey (CPS) data for the years 1988-90 and 2000-02, we then show that the level and dispersion of wages across occupations have changed substantially over the 1990s, 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 dispersion of wages across occupations. This evidence suggests that changes in occupational wage setting are a promising way of accounting for the U-shaped feature of changes in the wage distribution (Autor, Katz, and Kearney, 2006, 2008).

Second, we explicitly quantify the contribution of occupations, as summarized by the task content of jobs, in overall changes in the distribution of wages over the last three decades. We do so using a decomposition method based on the recentered influence function regression approach of Firpo, Fortin, and Lemieux (2009, 2011). This approach enables us to evaluate the contribution of changes in the returns to occupational tasks compared to other explanations such as de-unionization and changes in the labor market wide returns to general skills (labor market experience and education). We find that technological change and de-unionization played a relatively central role in the 1980s and 1990s, but had little effect in the 2000s. Increasing returns to education played an important role in all three decades, while offshorability became an important factor in the 1990s and, especially, in the 2000s.

The paper is organized as follows. In Section I, 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. Section II describes the wage data used, introduces the measures of task content computed from the O\*NET data, and explains how they are linked to the concepts of technological change and offshorability. Section III documents the changes in the level and dispersion of wages across occupations and shows that they are connected to our measures of the task content of jobs. The second part of the paper begins, in Section IV, with a short exposition of the decomposition methodology based on recentered influence function regressions. The ensuing results are presented in Section V and we conclude in Section VI.

## I. Wage Setting in Occupations

### A. 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 (2011) 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 observed over the last few decades. This is discussed in detail in Acemoglu and Autor (2011) 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, such as the “polarization” of the wage distribution of the 1980s. Second, the model does not provide insight on the contribution of occupations to changes in the wage structure because it does not draw any distinction between “skills” and “tasks”. Acemoglu and Autor (2011) 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 (2011), we go one step further here by allowing wages to vary across occupations *conditional* on the skills of workers, as in the 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. In Welch (1969), the wage  $w_{it}$  of worker  $i$  at time  $t$  is set as follows:

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

where the  $S_{ik}$ 's (for  $k = 1, \dots, K$ ) are skill components 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  $\theta_t$  is a base payment that a worker receives regardless of her skills.

As first pointed out by Rosen (1983), Welch (1969)'s model where wages solely depend on the bundle of skills supplied by each worker is unlikely to hold when workers are allocated to different tasks or occupations. Consider what happens when workers have the choice between several occupations that have different production functions (or skill requirements). Following Rosen (1983) and Acemoglu and Autor (2011), assume that the production function in an occupation simply depends on the sum of skills supplied by all workers in the occupation. Then, for each skill  $k$ , the aggregate supply of skill in the occupation is the sum of  $S_{ik}$  over all workers  $i$  in the occupation. Rosen (1983) shows that returns to skill only get 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 soundly 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 for the overall wage structure of differences and changes in wage setting across occupations.<sup>5</sup>

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

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

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  $\theta_{jt}$  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 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

<sup>5</sup>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.

face-to-face meetings with customers are important (e.g. sale managers). In occupations where face-to-face meetings are not essential (e.g. computer programmers), however, the returns to these skills have likely declined 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  $r_{jkt}$ .

### *B. Empirical test of the occupational wage setting model*

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 to the overall changes in the wage structure by computing counterfactual distributions based on alternative measures of  $r_{jkt}$  (and  $\theta_{jt}$ ). 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 consists, therefore, in seeing whether there are significant differences in the changes in both the level and 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, discussed in the next section, economists get a high score on the task “analyzing data or information”. Presumably, having good cognitive skills is quite important for performing these types of tasks. In the 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 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 summarize, 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 summarizing a large set of occupation dummies using a more limited number of tasks performed in these occupations.

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 flexible enough to allow for different changes in different parts of the distribution.

We do so by first estimating linear regression models that relate the changes in wages at different quantiles  $q$  of the wage distribution for each occupation,  $\Delta w_j^q$ , to the corresponding wage quantile in the base period,  $w_{j0}^q$ :

$$(3) \quad \Delta w_j^q = a_j + b_j w_{j0}^q + \lambda^q + \varepsilon_j^q,$$

where  $\lambda^q$  is a percentile-specific error component, which represents a generic change in the return to unobservable skills of the type considered by Juhn, Murphy, and Pierce (1993), and  $\varepsilon_j^q$  is an idiosyncratic error term. As we explain in more detail in the technical appendix, the occupation-specific intercepts ( $a_j$ ) and slopes ( $b_j$ ) from these regressions are two summary statistics that can be directly linked to changes in returns to skill requirements in each occupation,  $\Delta r_{jk}$ .<sup>6</sup> In addition, the intercept depends on changes in occupational wage differentials that are unrelated to skills,  $\Delta\theta_j$ . We use these parameters ( $a_j$  and  $b_j$ ) to characterize the changes in the wage distribution for each occupation. The intercepts  $a_j$  capture between-occupation changes in wage dispersion, while the slopes  $b_j$  capture within-occupation changes. As in the single index model of Card and Lemieux (1996),  $b_j > 0$  indicates an increase in returns to skill, while  $b_j < 0$  indicates the opposite.

In the second step of the analysis, we link the estimated intercepts and slopes ( $\hat{a}_j$  and  $\hat{b}_j$ ) to

<sup>6</sup>More precisely, we show that  $a_j = \Delta\theta_j + \sum_{k=1}^K \Delta r_{jk} \bar{\sigma}_{jk}$  and  $b_j \approx 1/\sigma_{j0}^2 [\sum_{k=1}^K (r_{jk0} \Delta r_{jk}) \cdot \sigma_{kj}^2 + 1/2 \Delta\sigma_{uj}^2]$ , where  $\sigma_{.j}$  are the related within-occupation standard deviations.

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  $h = 1, \dots, 5$ .

The second step regressions are

$$(4) \quad \hat{a}_j = \gamma_0 + \sum_{h=1}^5 \gamma_{jh} TC_{jh} + \mu_j,$$

and

$$(5) \quad \hat{b}_j = \delta_0 + \sum_{h=1}^5 \delta_{jh} TC_{jh} + \nu_j.$$

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, for occupations scoring high in terms of the routine aspect of the work performed, we should observe negative estimates of both  $a_j$  and  $b_j$ . Similarly, for occupations that involve face-to-face meetings are less likely to be offshored, we should observe positive estimates of  $a_j$  or  $b_j$  parameters.

## II. Data

### A. Wage Data

The empirical analysis is based on data for men from the Outgoing Rotation Group (ORG) Supplements of the Current Population Survey.<sup>7</sup> 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. At the beginning and end of the three decades we analyze, we pool several years of data together to improve the precision of the estimates. For the first period being analyzed (1976-78 to 1988-90), we start with data from the May CPS for the years 1976 to 1978.<sup>8</sup> For the second, and main period of analysis, we use 1988-90 as the base year and

<sup>7</sup>The data files were processed as in Lemieux (2006b) who provides detailed information on the relevant data issues.

<sup>8</sup>The reason we use the May CPS instead of the MORG CPS for 1979 or 1980 is that union status was not asked in MORG CPS until 1983. Since inequality was relatively stable during the 1970s (see DiNardo, Fortin, and Lemieux, 1996),

2000-02 as the end year to make sure we fully capture all the changes that occurred during the 1990s.<sup>9</sup> The base year for the final period is 2003-2004, which does not overlap with the end year of the 1988-90 to 2000-02 period. The reason for this discrepancy is that there was a major change in the coding of occupations when the 2000 census classification was introduced in 2003. For the sake of consistency in the coding of occupations, we only look at post-2002 data when performing the decomposition for recent years. This explains why the final period we consider goes from 2003-04 to 2009-10.<sup>10</sup>

Consistent with Autor, Katz and Kearney (2006), Figure 1 shows that 1988-90 to 2000-02 changes in real wages (\$1979) at each percentile of the wage distribution follow a U-shaped curve. In the figure, we also contrast these wage changes with those that occurred before (1976-78 to 1988-90) and after (2003-04 to 2009-10) the 1990s. The figure illustrates that wage changes in the top half of the distribution were quite similar during all time periods, though the changes have been more modest since 2003. 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 increased rapidly during the 1980s, but decreased sharply after 1988-90 as wages at the bottom grew substantially more than those in the middle of the distribution. The bottom part of the distribution has remained more or less unchanged since 2003. This is a bit surprising since recessions are typically believed to have a particularly negative impact at the bottom end of the distribution. More generally, wage changes for 2003-04 to 2009-2010 should be interpreted with caution since macroeconomic circumstances were very different during these two time periods. By contrast, the overall state of the labor market was more or less comparable in the other years considered in the analysis.<sup>11</sup>

### *B. Occupational Measures of Technological Change and Offshoring Potential*

Like many recent papers (Goos and Manning (2007), Goos, Manning and Salomons (2010), Crinó (2010)) that study the task content of jobs, and in particular their offshorability, we use

the precise choice of base year for studying changes in inequality during the 1980s should not have much impact on the results.

<sup>9</sup>By pooling three years of data at each end of the sample period, we obtain relatively large samples both in 1988-90 (268,497 observations) and 2000-02 (252,397 observations).

<sup>10</sup>Note that there was also a major change in the coding of occupations classification when the 1980 census classification was introduced in 1983. The timing is unfortunate as it coincides with the time period when inequality expanded the most dramatically (1980 to 1984). Since changes in occupational classification were not as dramatic as those that happened in the early 2000s, we used a crosswalk to keep a reasonably consistent definition of occupations between 1976-78 and 1988-90.

<sup>11</sup>The average unemployment rate for the 1976-78, 1988-90, 2000-02, and 2003-04 period is 6.2, 5.9, 4.8 and 5.8 percent, respectively, compared to 9.5 percent for 2009-10.

the O\*NET data to compute our measures of technological change and offshoring potential.<sup>12</sup> 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.<sup>13</sup> Our construction of an index of potential offshorability follows the pioneering work of Jensen and Kletzer (2010) [JK hereinafter] while incorporating some of the criticisms of Blinder (2007). The main concern of Blinder (2007) is the inability of objective indexes to take into account two important criteria for non-offshorability: a) 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 measures. The O\*NET content model organizes 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 and also 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 offshorability (and technology), and “face-to-face”, “on-site job”, and “decision-making”, thought to be negatively related to offshorability.<sup>14</sup> Our first category “infor-

<sup>12</sup> 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 (2010). Alternatively, using the German IAB/BIBB survey, Spitz-Oener (2006) and Dustmann, Ludsteck and Schönberg (2009) use task measures collected at the worker level.

<sup>13</sup> Blinder (2010) did not compute his index for Category IV occupations (533 occupations out of 817), that are deemed impossible to offshore. Although, Jensen and Kletzer (2010) report their index for 457 occupations, it is not available for many blue-collar occupations (occupations SOC 439199 and up).

<sup>14</sup> Appendix Table A2 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).

mation 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. Our second category “automation/routinization” is constructed using some work context measures to reflect the degree of potential automation of jobs and is an update on 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”. 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”.

For each occupation, the O\*NET provides information on the “importance” and “level” of required work activity and on the frequency of five categorical levels of work context.<sup>15</sup> 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

<sup>15</sup>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.

category  $h$  is, thus, computed as

$$(6) \quad TC_{jh} = \sum_{k=1}^{A_h} I_{jk}^{2/3} L_{jk}^{1/3} + \sum_{l=1}^{C_h} F_{jl} * V_{jl},$$

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

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.

Table 1 shows a number of summary statistics for the five normalized measures of task content.<sup>16</sup> The table reports the average value of the measures of task content for five major occupational groups. In Panel A, these broad occupations are constructed using the 1980-1990 Census occupation codes. Corresponding measures based on the 2002 census occupation codes are reported in Panel B. Since most of the empirical analysis presented below focuses on the 1990s, we limit our discussion to the results reported in Panel A.

The results reported in Table 1 are generally consistent with the evidence reported in related studies. Professional, managerial and technical occupations have the highest score in terms of their use of information technology, and the lowest score for automation. As a result, these high wage occupations are likely to benefit the most from technological change. Interestingly, this broad occupation group also gets the highest score in terms of face-to-face interactions and decision making, suggesting that they should not be too adversely affected by offshoring.<sup>17</sup> At the other end of the spectrum, production workers and operators have a relatively low score in terms of their use of information technology and a high score for automation. These jobs also involve little face-to-face interactions or decision making. Therefore, both offshoring and technological change are expected to have an adverse impact on wages in these occupations.

The pattern of results for on-site work is more complex. Consistent with our expectations, primary, construction and transport workers have the highest score for on-site work, while clerical

<sup>16</sup>The range of these measures goes from zero to one since we normalize the task measures by dividing them by their maximum value observed over all occupation. These normalized tasks measures provide a useful ranking of occupations along each of these five dimensions, but the absolute values of the task measures have no particular meaning.

<sup>17</sup>Including decision-making in our offshorability measure thus allows a sharper distinction between managerial and clerical jobs.



and sales workers have the lowest score. Interestingly, production workers and operators also get a high score for working on-site, suggesting that they should not be too affected by offshoring. This illustrates the limits of the O\*NET measures as a way of capturing the offshorability of jobs. While it is true that production workers tend to work on a specific site, the whole production process could still be offshored. This is quite different from the case of construction workers for whom the “site” has to be in the United States. As a result, the effect of on-site work on wages should be interpreted with some caution.

### III. Occupational Wage Profiles: Results

In this section, we first estimate the linear regression models for within-occupation quantiles from equation (3), and then link the estimated slope and intercept parameters to our measures of task content from the O\*NET as in equations (4) and (5). We refer to these regression models as “occupation wage profiles”. We focus this first part of the analysis on the 1990s as it represents the time period when most of the polarization of wages documented by Autor, Katz and Kearney (2006) occurred.

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.<sup>18</sup> 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.<sup>19</sup> 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  $q = 10, 20, \dots, 90$ . Finally, all the regression estimates are weighted by the number of observations (weighted using the earnings weight from the CPS) in each occupation.

Detailed estimates of several specifications for equation (3) are presented in Appendix Table

<sup>18</sup>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).

<sup>19</sup>The bias could be adjusted using a measurement-error corrected regression approach, as in Card and Lemieux (1996), or an instrumental variables approach.

A3 and discussed in the Technical Appendix. The main finding is that occupation-specific slopes and intercepts both have to be included in the regression models to adequately account for the observed wage changes. This general model explains over 90 percent of the variation in the data, and all of the curvature (or U-shape feature) that characterizes wage changes over that period.

We illustrate the fit of the model by plotting occupation-specific regressions for the 30 largest occupations curves in Figure 2.<sup>20</sup> While it is not possible to see what happens for 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 set of occupational wage profiles generally follow the U-shaped pattern observed in the raw data. In light of the discussion in Section 2, this suggests that skills that used to be valuable in low-wage occupations are less valuable than they used to be, while the opposite is happening in high-wage occupations.

We next explore this hypothesis more formally by estimating the regression models in equations (4) and (5) that link the intercept and slopes of the occupation wage change profiles to the task content of occupations.<sup>21</sup> The results are reported in Table 2. In the first four columns of Table 2, we include task measures separately in the regressions (one regression for each task measure). 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.

As some tasks involving the processing of information may be enhanced by ICT technologies, we would expect a positive relationship between our “information content” task measure and the measures of occupational wage changes. On the other hand, 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 occupational wage changes with high degree of “automation” to decline over time.<sup>22</sup> Although occupations in the middle of the wage distribution may be most vulnerable to technological change, some also involve relatively more “on-site” work (e.g. repairmen) and may, therefore, be less vulnerable to offshoring. We also expect workers in occupations with a high level of “face-to-face” contact, as well as those

<sup>20</sup>To avoid overloading the graph, we exclude ten occupations that account for the smallest share of the workforce (less than one percent of workers in each of these occupations).

<sup>21</sup>To be consistent with equation (A-8), 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.

<sup>22</sup>In Appendix Figure A1, we show the connection between task measures and average occupational wages for all three digit occupations. While some of the measures (information content and no decision making) tend to be monotonically related to occupation wages, others follow a more interesting shape. Consistent with Autor, Levy and Murnane (2003), the “automation” task follows an inverse U-shaped curve.

with a high level of “decision-making”, to do relatively well in the presence of offshoring.

The strongest and most robust result in Table 2 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 six of the eight specifications reported in Table 2. The other “technology” variable, information content, has generally a positive and significant effect on both the intercept and the slope, as expected, when included by itself in columns 1 to 4. 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 at the bottom end of the wage distribution. 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. As discussed earlier and argued by Crisculo and Garicano (2010), increasing the offshoring of complementary tasks may increase wages at the top end of the wage distribution.

The results reported in Table 2 generally conform to expectations. The effect of “no face to face” and “no decision making” is generally negative. By contrast, the effect of “no on-site work” is generally positive, which may indicate that on average we are capturing the positive effect of offshoring. Another 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).

More importantly, Table 2 shows that the task measures explain most of the variation in the slopes, though less 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.<sup>23</sup>

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 on the slope of the occupational wage profiles. Second, tasks account for a large

<sup>23</sup>Determining which tasks to include remains an important challenge of the approach. Here we simply follow the main tenants of the literature.

fraction 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 task measures can play a substantial role in explaining the U-shaped feature of the raw data illustrated in Figure 1. These results also show why models that do not account for within-occupation changes are likely to miss a significant part of the contribution of the task content of jobs to changes in the wage distribution.

Although the analysis presented above helps illustrate the mechanisms through which occupations play a role in changes in the wage structure, it does not precisely quantify the relative contribution of occupational factors to these changes.<sup>24</sup> We next examine the explanatory power of occupational tasks in the context of a formal decomposition of changes in the wage distribution.

#### IV. Decomposing Changes in Distributions Using RIF-Regressions

In this section, we show how to formally decompose changes in the distribution of wages into the contribution of occupational and other factors using the recentered influence function (RIF) regression approach introduced by Firpo, Fortin, and Lemieux (2009).<sup>25</sup> As is well known, a standard regression can be used to perform a Oaxaca-Blinder decomposition for the mean of a distribution. RIF-regressions allow us to perform the same kind of decomposition for any distributional parameter, including percentiles.

In general, any distributional parameter can be written as a functional  $\nu(F_Y)$  of the cumulative distribution of wages,  $F_Y(Y)$ .<sup>26</sup> 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-Blinder decomposition, the wage structure effect only depends on changes in the conditional mean of wages,  $\mathbb{E}(Y|X)$ . 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 framework. We focus on differences in the wage distributions for two time periods, 1 and 0. For a worker  $i$ ,

<sup>24</sup>Two limitations of the approach are the linearity of the specification and the focus on the occupation-specific distribution of wages. Also the approach does generally control for other commonly considered factors, such as education, experience, unionization, etc.

<sup>25</sup>Firpo, Fortin, and Lemieux (2007, 2011) explain in more detail how to perform these decompositions, and show how to compute the standard errors for each element of the distribution. Here, we simply present a short summary of the methodology.

<sup>26</sup>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.

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. Therefore, for each  $i$  we can define the observed wage,  $Y_i$ , as  $Y_i = Y_{1i} \cdot T_i + Y_{0i} \cdot (1 - T_i)$ , where  $T_i = 1$  if individual  $i$  is observed in period 1, and  $T_i = 0$  if individual  $i$  is observed in period 0.<sup>27</sup> There is also a vector of covariates  $X \in \mathcal{X} \subset \mathbb{R}^K$  observed in both periods.

Consider  $\Delta_O^\nu$ , the overall change over time in the distributional statistic  $\nu$ . We have

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

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

Estimating this type of 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 distribution under the ignorability assumption.

However, the main goal of this paper is to separate the contribution of different subsets of covariates to  $\Delta_O^\nu$ ,  $\Delta_S^\nu$ , and  $\Delta_X^\nu$ . 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. 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  $\text{RIF}(y; \nu)$  of the statistic  $\nu$ . The recentering consists of adding back the distributional statistic  $\nu$  to the influence function  $\text{IF}(y; \nu)$ :  $\text{RIF}(y; \nu) = \nu + \text{IF}(y; \nu)$ . Note that in the case of the mean where the influence function is  $\text{IF}(y; \mu) = y - \mu$ , we have  $\text{RIF}(y; \mu) = \mu + (y - \mu) = y$ . Since the  $\text{RIF}(y; \mu)$  is simply the outcome variable  $y$ , the RIF-regression for the mean corresponds to a standard wage regression.

It is also possible to compute the influence function for many other distributional statistics. Of particular interest is the case of quantiles. The  $\tau$ -th quantile of the distribution  $F$  is defined

<sup>27</sup>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.

as the functional,  $Q(F, \tau) = \inf\{y | F(y) \geq \tau\}$ , or as  $q_\tau$  for short. Its influence function is:

$$\text{IF}(y; q_\tau) = \frac{\tau - \mathbb{1}\{y \leq q_\tau\}}{f_Y(q_\tau)}.$$

The recentered influence function of the  $\tau^{\text{th}}$  quantile is  $\text{RIF}(y; q_\tau) = q_\tau + \text{IF}(y; q_\tau)$ .

Consider  $\gamma_t^\nu$ , the estimated coefficients from a regression of  $\text{RIF}(y_t; \nu)$  on  $X$

$$\gamma_t^\nu = (\mathbb{E}[X \cdot X^\top | T = t])^{-1} \cdot \mathbb{E}[\text{RIF}(y_t; \nu) \cdot X | T = t], \quad t = 0, 1.$$

Because of the law of iterated expectations, distributional statistics can be expressed in terms of expectations of the conditional recentered influence functions,

$$\nu(F_t) = \mathbb{E}_X [\mathbb{E}[\text{RIF}(y_t; \nu) | X = x]] = \mathbb{E}[X | T = t] \cdot \gamma_t^\nu.$$

In particular, the  $\tau^{\text{th}}$  quantile RIF-regression aggregates to the unconditional quantile of interest and allows us to capture both the between and the within effects of the explanatory variables.

By analogy with the Oaxaca-Blinder decomposition, we could write the wage structure and composition effects as:

$$\Delta_S^\nu = \mathbb{E}[X | T = 1]^\top (\gamma_1^\nu - \gamma_0^\nu) \quad \text{and} \quad \Delta_X^\nu = (\mathbb{E}[X | T = 1] - \mathbb{E}[X | T = 0])^\top \gamma_0^\nu.$$

This particular decomposition is very easy to compute since it is similar to a standard Oaxaca-Blinder 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, 2011) suggest a solution to this problem based on an hybrid approach that involves both reweighting and RIF-regressions. 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. This means that  $\gamma_1^\nu$  and  $\gamma_0^\nu$  may be different just because they are estimated for different distributions of  $X$  even if the wage structure remains unchanged over time.

But reweighting will adjust for this problem. Letting  $\Psi(X)$  be the reweighing function,

$$\Psi(X) = \frac{\Pr(T = 1|X)/\Pr(T = 1)}{\Pr(T = 0|X)/\Pr(T = 0)}.$$

that makes the distributions of  $X$ 's in period 0 similar to that of period 1, one can estimate the counterfactual mean as  $\bar{X}_{01} = \sum_{i \in 0} \widehat{\Psi}(X_i) \cdot X_i \rightarrow \bar{X}_1$ , and the counterfactual coefficients  $\widehat{\gamma}_{01}^\nu$  as the coefficients from a regression of  $\widehat{\text{RIF}}(Y_0; \nu)$  on the reweighted sample  $\{X_0; \widehat{\Psi}(X_0)\}$ .<sup>28</sup> Then the difference  $\widehat{\gamma}_1^\nu - \widehat{\gamma}_{01}^\nu$  reflects a true change in the wage structure.

The composition effect  $\widehat{\Delta}_{X,R}^\nu$  can be divided into a pure composition effect  $\widehat{\Delta}_{X,p}^\nu$  using the wage structure of period 0 and a component measuring the specification error,  $\widehat{\Delta}_{X,e}^\nu$ :

$$\begin{aligned} \widehat{\Delta}_{X,R}^\nu &= (\bar{X}_{01} - \bar{X}_0) \widehat{\gamma}_0^\nu + \bar{X}_{01} [\widehat{\gamma}_{01}^\nu - \widehat{\gamma}_0^\nu]. \\ (7) \qquad &= \widehat{\Delta}_{X,p}^\nu + \widehat{\Delta}_{X,e}^\nu \end{aligned}$$

Similarly, the wage structure effect can be written as

$$\begin{aligned} \widehat{\Delta}_{S,R}^\nu &= \bar{X}_1 (\widehat{\gamma}_1^\nu - \widehat{\gamma}_{01}^\nu) + (\bar{X}_1 - \bar{X}_{01}) \widehat{\gamma}_{01}^\nu \\ (8) \qquad &= \widehat{\Delta}_{S,p}^\nu + \widehat{\Delta}_{S,e}^\nu \end{aligned}$$

and reduces to the first term  $\widehat{\Delta}_{S,p}^\nu$  given that the reweighting error  $\widehat{\Delta}_{S,e}^\nu$  goes to zero as  $\bar{X}_{01} \rightarrow \bar{X}_1$ .

Again, this decomposition is very easy to compute as it corresponds to two standard Oaxaca-Blinder decompositions performed on the estimated recentered influence functions. The first compares time period 0 and the reweighted time period 0 that mimics time period 1 and allows us to obtain the pure composition effects. The second compares the time period 1 and the

<sup>28</sup>The reweighting function is computed as the ratio of the predicted probabilities obtained from a logit specification that includes a rich set of interaction between the explanatory variables.

reweighted time period 0, and allows use to obtain the pure wage structure effects.

## V. Decomposition Results: Occupational Characteristics vs. Other Factors

The covariates included in the regressions reflects the different explanations that have been suggested for the changes in the wage distribution over our sample period. 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 below. We also include controls for marital status and race in all the estimated models.<sup>29</sup>

Before showing the decomposition results, it is useful to discuss some features of the estimated RIF-coefficients across the different wage quantiles.<sup>30</sup> For example, the effect of the union status across the different quantiles is highly non-monotonic. In both 1988-90 and 2000-2002, the effect first increases up to 0.4 around the median, and then declines (Appendix Figure A2). This indicates that unions *increase* inequality in the lower end of the distribution, but *decrease* inequality even more in the higher end of the distribution. The results for unions illustrate an important feature of RIF-regressions for quantiles, namely that they capture the effect of covariates on both the between- and within-group components 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, we find that three of our five task measures have non-monotonic impact across the different percentiles of the wage distribution. Both “information” and “no face-to-face” have an inverse U-shaped impact, while “automation” has a largely negative U-shaped impact. Furthermore, changes over time in the effect of these first two task measures shows a declining effect in the lower middle of the distribution, but an increasing effect in the upper middle of the distribution. Changes over time in the wage effect of “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 technologies. Changes

<sup>29</sup>The sample means for all these variables are provided in Appendix Table A1.

<sup>30</sup>The RIF-regression coefficients for the 10<sup>th</sup>, 50<sup>th</sup>, and 90<sup>th</sup> quantiles in 1988-90 and 2000-02, along with their (robust) standard errors are reported in Appendix Table A4. The RIF-regression coefficients for the variance and the Gini are reported in Appendix Table A5. We also plot in Appendix Figure A2 (standard covariates) and Appendix Figure A3 (five task measures) the estimated coefficients from RIF-regressions for 19 different wage quantiles (from the 5<sup>th</sup> to the 95<sup>th</sup> quantile) equally spread over the whole wage distribution.



over time in the impact of the other tasks measures appear less important.

The results of the decomposition are presented in Figures 3-5. Tables 3 and 4 also summarize the results for the standard measures 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 some college, 15 to 19 years of potential experience, and occupational task measures at half a standard deviation below their sample averages.<sup>31</sup> A richer specification with additional interaction terms is used to estimate the logit models used in the computation of the reweighting factor.<sup>32</sup> The reweighting approach performs well in the sense that the reweighted means of the covariates for the base period are very close to those for the end period.<sup>33</sup>

As is well known (e.g. Oaxaca and Ransom, 1999), the detailed wage structure part of the decomposition (equation (8)) 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.<sup>34</sup> Here, we normalize the task measure variables such that the average difference between the end and beginning period is equal to half a standard deviation of the raw measure. The wage structure effect for each task measure can be interpreted as the change over time in the wage impact of a half a standard deviation increase in the measure.<sup>35</sup> We also note that any “composition” effect associated with the task measures are linked to changes in the shares of occupations over time, as our measures of task requirements for each occupation are invariant over time.

<sup>31</sup>We use “some college” as the base group as it represents the modal education group in the 1990s and 2000s. For the 1976-78 to 1988-90 period, we use high school graduates as the base group as it was still the modal education group during that period.

<sup>32</sup>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.

<sup>33</sup>The reweighting error is the second term in equation (8). If the reweighting was replicating the means perfectly, we would have  $\bar{X}_1 = \bar{X}_{01}$  and the reweighting error would be equal to zero. In Appendix Figure A5, the reweighting error corresponds to the difference between the total composition effect obtained by reweighting and with the RIF-regressions and is found to be very small and not significant.

<sup>34</sup>As discussed in Fortin, Lemieux and Firpo (2011), automatic normalization solutions to this issue are not satisfactory, rather the choice of a reasonable and interpretable base group is preferred.

<sup>35</sup>The choice of half of a standard deviation is based on the following reasoning. In Table 1, the difference between the mean value of task measures for all occupations and the mean for the major group with lowest mean ranges from 45 percent to 101 percent of the standard deviation. For example, the mean for automation in Panel A is 0.6871, which is 0.1014 (or 0.77 standard deviations) above the mean for professional, managerial, and technical occupations. This suggests that occupations at half a standard deviation below the mean are reasonably representative of a large group of occupations with relatively low values of the task measures. Thus, we use this criteria as a uniform way of choosing the base group for all five tasks measures.

### A. Overall Decomposition Results

Figure 3 shows the overall change in (real log) wages at each percentile  $\tau$ ,  $(\Delta_{\tau}^w)$ , and decomposes this overall change into a composition  $(\Delta_X^{\tau})$  and wage structure  $(\Delta_S^{\tau})$  effect.<sup>36</sup> Consistent with Autor, Katz and Kearney (2006), Figure 3b shows that the overall change between 1988-90 and 2000-02 is U-shaped as wage dispersion increases in the top end but declines in the lower end of the distribution. This stands in sharp contrast with the situation that prevailed in the early 1980s. Figure 3a shows that the corresponding curve is positively sloped for all quantiles as wage dispersion increases at all points of the distribution (as in Juhn, Murphy, and Pierce, 1993). Figure 3c shows that although the wage distribution has been much more stable in recent years, there is still a modest increase in inequality during the 2003-04 to 2009-10 period. On the one hand, this is not surprising since the period under consideration is half as long as the two other periods considered in Figure 3. On the other hand, one could have expected a more dramatic drop in wages at the bottom end of the distribution given the adverse macroeconomic conditions of the last few years.

Table 3 summarizes the changes reported in Figure 3 using a few standard measures of wage dispersion. There is a large increase in inequality measures, such as the variance and the 90-10 gap, that capture wage changes over the entire distribution in the 1980s, and a more modest increase in later periods. Consistent with Figure 3, inequality at the top end of the distribution (the 90-50 gap) increases in all time periods. By contrast, the 50-10 gap increases before 1990 and after 2003, but declines substantially during the 1990s.

Figure 3 and Table 3 also show that, consistent with Lemieux (2006b), composition effects have contributed to a substantial increase in inequality since the late 1970s. For instance, composition effects account for between 20 percent and 45 percent of the growth in the 90-50 gap in each of the three time periods. Looking at cumulative changes over all time periods, composition effects account for all of the change in the 50-10 gap, and about a third of the change in the 90-50 gap.<sup>37</sup> But while composition effects account for a sizable part of the growth in overall inequality, it fails to explain the U-shape pattern observed during the 1990s. As a result, all of the 1990s U-shape feature in the change in the wage distribution is captured by the wage structure effect.

<sup>36</sup>The composition effect reported in Figure 3 only captures the component,  $\hat{\Delta}_{X,p}^{\nu}$  from equation (8). The specification error,  $\hat{\Delta}_{X,e}^{\nu}$ , corresponds to the difference between the total composition effect obtained by reweighting and RIF-regression methods illustrated in Appendix Figure A4. The figure shows that RIF-regressions capture quite accurately the overall trend in composition effects, though there are a number of small discrepancies at various points along the wage distribution, likely reflecting spikes in the wage distribution.

<sup>37</sup>The total change in the 50-10 and 90-50 gap between 1976-78 and 2009-10 is 0.0363 and 0.2210, respectively. The corresponding composition effect is 0.0353 and 0.0770, respectively.

### B. Detailed Decomposition Results

Figure 4 moves to the next step of the decomposition using RIF-regressions to apportion the composition effect to the contribution of each set of covariates. Figure 5 does the same for the wage structure effect. To simplify the presentation of the results, Figure 4 reports the composition effect for five set of explanatory factors: union status, education, experience, offshorability and technological change.<sup>38</sup> The effect of the other covariates used in the RIF-regressions (race and marital status) is generally small. We report it in Panel A of Table 4 under the “other” category. For the sake of simplicity, we focus the discussion on the impact of each factor in the lower and upper part of the distribution. Those are also summarized in terms of the 50-10 and 90-50 gaps for the main analysis period (1988-90 to 2000-02) in Table 4.

First consider composition effects for the 1988-90 to 2000-02 period. 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 0.033 and -0.002 for the 50-10 and 90-50 gaps, respectively. Composition effects linked to factors other than unions thus go in the “wrong direction” as they account for rising inequality at the bottom end while inequality is actually rising at the top end of the distribution.

In contrast, composition effects linked to unions (the impact of de-unionization) reduce inequality at the low end (effect of -0.013 on the 50-10 gap) and increase inequality at the top end (effect of 0.027 on the 90-50). Note that, as in a Oaxaca-Blinder decomposition, these effects on the 50-10 and the 90-50 gap can be obtained directly by multiplying the 5.3 percent decline in the unionization rate (Appendix Table A1) by the RIF-regression estimates of the union effects for 1988-90 (Appendix Table A4). The resulting effect of de-unionization accounts for 24 percent of the total change in the 50-10 gap, and 30 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).

The results for the 1976-78 to 1988-90 period are reported in Figure 4a. Since unionization declined more dramatically during in the 1980s (9.3 percentage point decline) than in the 1990s (5.3 percentage point decline), the estimated contribution of de-unionization to inequality changes is also larger during the earlier period. As in the 1990s, de-unionization has a larger and positive

<sup>38</sup>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.

effect on inequality growth at the top end, and a smaller negative impact at the low end for the distribution. Interestingly, the results reported in Figure 4a are very similar to those obtained by DiNardo, Fortin, and Lemieux (1996) using a different method.<sup>39</sup> The effect of the other factors is relatively modest and does not account for much of the large increase in inequality observed during that period.

The rate of unionization only declined by 0.5 percentage points after 2002 (see Appendix Table A1). As a result, composition effects linked to de-unionization are negligible for the 2003-04 to 2009-10 period. Changes in the distribution of educational attainment has a small impact on inequality growth, while the effect of the other factors is negligible.

Figure 5 reports the corresponding estimates for the wage structure effect. As in the case of composition effects, the contribution of each set of covariates to the wage structure effect for 1988-90 to 2000-02 is reported in Panel B of Table 4. Table 4 (and Appendix Figure A6) also reports the change in the intercept (constant) in the RIF-regressions. As in a standard Oaxaca-Blinder decomposition, the change in intercepts captures the part of the wage structure effect that cannot be accounted for by the covariates.<sup>40</sup> By definition, the change in intercepts represents the change in the wage distribution for the base group used in the RIF-regression analysis. One can therefore interpret that component of the decomposition as the residual (or within-group) change for the base group.

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 4 for the 1988-90 to 2000-02 periods. The results indicate that -0.050 of the -0.080 change (decline) in the 50-10 gap due to wage structure effects remains unexplained (the effect of the “constant” in Table 4). By contrast, changes in the return to covariates account for all (and even more) of the 0.057 change in the 90-50 gap linked to the wage structure, as the residual term (the constant) is now negative. Taken at face value, these results suggest a notable decline in residual wage inequality at most points of the distribution (see also Appendix Figure 6b). This finding is consistent with Lemieux (2006b) who also finds that the (composition adjusted) residual variance declined over that period.

Focusing on the contribution of the different covariates, Table 4 shows that changes in the wage structure linked to education play a substantial role at the top end of the distribution, but do not

<sup>39</sup>Using a conditional reweighting approach, DiNardo, Fortin, and Lemieux find that de-unionization reduced the 50-10 gap by 0.019 and increased the 90-50 gap by 0.040 between 1979 and 1988. The corresponding numbers from Figure 5a are 0.015 and 0.034, respectively.

<sup>40</sup>More formally, the total wage structure effect,  $\widehat{\Delta}_{S,p}^{\tau}$ , is the sum of the component explained by the RIF-regression models,  $\sum_{k=2}^M \overline{X}_1 (\widehat{\gamma}_1^{\nu} - \widehat{\gamma}_{01}^{\nu})$ , and the residual component  $\widehat{\gamma}_{1,1}^{\nu} - \widehat{\gamma}_{01,1}^{\nu}$  captured by the change in the intercepts.

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 a decline in the returns to experience 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 4 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. This can also be seen in Figure 5b where the wage structure effects linked to technology and offshoring both follow a distinct U-shape that closely mirrors the shape of the overall change in the wage distribution (Figure 4b). The important finding here is that technological change and offshorability, as captured by the occupations task measures included in the RIF-regression, go a long way towards explaining the polarization of wages observed in the 1990s.

This finding also raises an interesting question. If technological change and offshorability are indeed contributors to the polarization of wages during the 1990s, should they not also matter in other time periods? Figure 5a shows that, as in the 1990s, technological change had a large inequality enhancing effect at the top end of the distribution during the 1980s. In the 1980s, however, technological change also had a large impact at the bottom end of the distribution. This is consistent with the view that technological change was skill-biased during the 1980s (e.g. Berman, Bound, and Griliches, 1994), but "routine-biased" during the 1990s (Autor, Levy, and Murnane, 2003). By contrast, offshorability has little impact during the 1980s, which is consistent with service offshoring being a relatively recent phenomenon, made possible by the development of the internet and related technologies.

The other elements of the wage structure decomposition for 1976-78 to 1988-90 are generally consistent with the previous literature. For instance, Figure 5a shows that the growth in returns to education and experience contribute to the increase in inequality over that period. Interestingly, little residual growth in inequality (the "effect" of the constant) is left after all the other factors have been accounted for (Appendix Figure 6a). This may sound surprising since Juhn, Murphy, and Pierce (1993) and others show that residual inequality increased substantially during the 1980s. The difference here is that we add occupational task measures in the decomposition exercise. This suggests that technological change, as captured by the changing returns to the task measures, captures most of the residual change left after controlling for standard variables

like experience and education.

Finally, Figure 5c shows that, as in the 1990s, offshorability has contributed to the polarization of wages since the early 2000s. The fact that offshorability has a substantial effect on wages in recent years is consistent with the view of many labor market observers who have stressed the importance of offshoring, as opposed to technological change, in recent changes in the U.S. labor market.<sup>41</sup> Interestingly, Figure 5c also shows that technological change only has a modest effect on wage changes between 2003-04 to 2009-10, suggesting that the computer/internet revolution may have run its course in terms of labor market impacts.

While the detailed decompositions presented in this paper capture most of the explanatory factors suggested in the inequality literature, they do not account for the minimum wage which played an important role at the bottom end of the wage distribution during the 1980s (DiNardo, Lemieux, and Fortin, 1996, and Lee, 1999). As a crude way of capturing the impact of the minimum wage, we include in the RIF-regressions the fraction of workers potentially affected by changes in the minimum wage for each age and education group.<sup>42</sup> Since the real value of the minimum wage was relatively stable from 1988-90 and the later periods considered here, we only report the results of this exercise for the 1976-78 to 1988-90 when the minimum wage declined dramatically in real terms. The results (reported in Appendix Figure A6) show that the decline in the minimum wage had a large impact on the 5th and 10th percentiles of the wage distribution, but little effect for higher quantiles. The effect at the bottom end is substantial and accounts for 0.043 of the 0.070 total increase in the 50-10 gap during this period. Thus, except for the very bottom of the distribution, introducing the minimum wage does not change the substantive findings for the 1976-78 to 1988-90 period discussed above.

A number of interesting conclusions emerge from these detailed wage decompositions. First, the composition effect linked to de-unionization accounts for a substantial part of the change in inequality both at the lower (50-10) and upper (90-50) end of the distribution during the 1990s. Second, the changing wage structure effects linked to unionization, education, and the occupational task measures of technology and offshorability all help account for the changing wage distribution during the 1980s and 1990s. One important contribution of the paper is to show that occupational task measures substantially enhance the explanatory power of our

<sup>41</sup>See Blinder (2007) and the references therein.

<sup>42</sup>For the period 1976-78 to 1988-90 (the only one considered here), we compute the fraction of workers in 1988-90 who earn more than the minimum wage in those years, but less than the (average) real value of the minimum wage for 1976-78. We define workers in this wage range as those potentially affected by a change of the minimum wage since their wage would have likely been higher if the minimum wage had remained constant in real terms.

decomposition exercises. Furthermore, the pattern of results is consistent with the view that technological change was skill-biased during the 1980s, “routine-biased” during the 1990s, but no longer played much of a role in the years 2000. By contrast, offshorability only started playing an important role in the 1990s, and has contributed to the polarization of wages in both the 1990s and 2000s.

## VI. 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 of changes in within-occupation wage percentiles between 1988-90 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 offshorability) 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 Firpo, Fortin, and Lemieux (2009). The results indicate that technological change and offshorability are two factors that help enhance our ability to account for the observed changes in the distribution of wages over the last three decades.

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 (2011), we conclude that it is essential to go beyond general 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

- ACEMOGLU, D. AND D.H. AUTOR (2011), "Skills, Tasks, and Technologies: Implications for Employment and Earnings," in O. Ashenfelter and D. Card, eds., *Handbook of Economics*, Amsterdam: North-Holland, Vol. IV.B.: 1043-1172.
- AUTOR, D.H. (2010) "The Polarization of Job Opportunities in the U.S. Labor Market Implications for Employment and Earnings," The Center for American Progress and The Hamilton Project.
- 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*, 90(2): 300–323.
- AUTOR, D.H., KATZ, L.F. AND M.S. KEARNEY (2006), "The Polarization of the U.S. Labor Market," *American Economic Review* 96(2): 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.
- BERMAN, E., J. BOUND, AND Z. GRILICHES (1994), "Changes in the Demand for Skilled Labor within U.S. Manufacturing: Evidence from the Annual Survey of Manufacturers", *Quarterly Journal of Economics*, 109(2): 367-397.
- 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.
- CARD, D. AND T. LEMIEUX (1996) "Wage Dispersion, Returns to Skill, and Black-White Wage Differentials" *Journal of Econometrics*, 74(2): 319-361.



- CRINÒ, R. (2010) “Service Offshoring and White-Collar Employment,” *Review of Economic Studies*, 77(2): 595-632.
- CRISCUOLO, C. AND L. GARICANO (2010) “Offshoring and Wage Inequality: Using Occupational Licensing as a Shifter of Offshoring Costs,” *American Economic Review*, 100(2): 439-443.
- DiNARDO, J., N.M. FORTIN, AND T. LEMIEUX (1996), “Labor Market Institutions and the Distribution of Wages, 1973-1992: A Semiparametric Approach,” *Econometrica*, 64(6): 1001-1044.
- DUSTMANN, C., J. LUDSTECK AND U. SCHÖNBERG (2009) “Revisiting the German Wage Structure,” *Quarterly Journal of Economics*, 124(2): 809–842.
- FIRPO, S. FORTIN, N.M., AND T. LEMIEUX (2007) “Decomposing Distribution Using Centered Influence Function Regressions,” Unpublished manuscript, PUC-Rio and UBC
- FIRPO, S. FORTIN, N.M., AND T. LEMIEUX (2009) “Unconditional Quantile Regressions,” *Econometrica* 77(3): 953-973.
- FEENSTRA, R, AND G. HANSON (2003), “Global Production Sharing and Inequality: A Survey of Trade and Wages,” In E.K. Choi and J. Harrigan (eds) *Handbook of International Trade*, Oxford: Blackwell, 146-185.
- FORTIN, N.M., T. LEMIEUX, AND S. FIRPO (2011) “Decomposition Methods in Economics,” in O. Ashenfelter and D. Card, eds., *Handbook of Economics*, Amsterdam: North-Holland, Vol. IV.A: 1-102
- 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.
- GIBBONS, R., L.F. KATZ, T. LEMIEUX, AND D. PARENT (2005) “Comparative Advantage, Learning, and Sectoral Wage Determination,” *Journal of Labor Economics* 23(4): 681-724.
- GOOS, M. AND A. MANNING (2007) “Lousy and Lovely Jobs: The Rising Polarization of Work in Britain”, *Review of Economics and Statistics*, 89(1): 118-133

- GOOS, M., A. MANNING AND A. SALOMONS (2009) “The Polarization of the European Labor Market”, *American Economic Review*, 99(2): 58-63.
- GOOS, M., A. MANNING, AND A. SALOMONS (2010) “Recent Changes in the European Employment Structure: The Roles of Technological Change, Globalization and Institutions,” unpublished manuscript.
- GROSSMAN, G. M., AND E. ROSSI-HANSBERG (2008) “Trading Tasks: A Simple Theory of Offshoring,” *American Economic Review*, 98(5): 1978—97.
- 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. B. AND L. G. KLETZER, (2010) “Measuring the Task Content of Offshorable Services Jobs, Tradable Services and Job Loss,” In *Labor in the New Economy*, Katharine Abraham, Mike Harper, and James Spletzer, eds., Chicago: University of Chicago Press. 309-335.
- 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.
- LEE, D.S. (1999) “Wage Inequality in the United States during the 1980s: Rising Dispersion or Falling Minimum Wage”, *Quarterly Journal of Economics* 114(3): 977-1023
- 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): 21-48.
- MURPHY, K.M., AND F. WELCH (1991) “The Role of International Trade in Wage Differentials,” in M. Koster (ed.) *Workers and Their Wages*, Washington DC: American Enterprise Institute Press, 39-69.
- OAXACA, R. AND M.R. RANSOM (1999), “Identification in Detailed Wage Decompositions,” *Review of Economics and Statistics* 81(1): 154–157.

- ROSEN, S. (1983), "Specialization and human capital" *Journal of Labor Economics* 1(1): 43–49.
- SPITZ-OENER, A. (2006) "Technical Change, Job Tasks, and Rising Educational Demands: Looking Outside the Wage Structure," *Journal of Labor Economics*, 24(2): 235-270.
- WELCH, F. (1969) "Linear Synthesis of Skill Distribution," *Journal of Human Resources*, 4(3): 311-327 .

Table 1. Average O\*NET Indexes by Major Occupation Group

O*NET Indexes	Technology		Not On-Site	Offshorability	
	Information	Automation		No Face-to-Face	No Decision Making
Panel A: using 1980-1990 Standard Occupation Codes					
Overall Mean	0.6845	0.6871	0.4072	0.3171	0.3105
Standard Deviation	0.1519	0.1313	0.2089	0.1062	0.1057
Professional, Managerial, Technical	0.8274	0.5857	0.5560	0.2498	0.2244
Clerical, Sales	0.7067	0.7177	0.6095	0.3037	0.3692
Production, Operators	0.6020	0.8205	0.2430	0.4197	0.3608
Primary, Construction, Transport	0.6075	0.6993	0.2029	0.3395	0.3149
Service	0.5549	0.6437	0.4724	0.2762	0.3747
Panel B: using 2002 Census Codes					
Overall Mean	0.6786	0.6310	0.4386	0.3155	0.3074
Standard Deviation	0.1188	0.1477	0.2084	0.1094	0.1083
Professional, Managerial, Technical	0.8126	0.5644	0.5729	0.2661	0.2285
Clerical, Sales	0.7166	0.6825	0.6363	0.3047	0.3601
Production, Operators	0.5892	0.7630	0.2752	0.4299	0.3709
Primary, Construction, Transport	0.5908	0.6469	0.2271	0.3394	0.3205
Service	0.5844	0.5843	0.4699	0.2566	0.3343

Note: The overall means are based on 510 basic occupations in Panel A and on 505 basic occupations in Panel B.

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

O*NET Indexes	Tasks Entered Separately				Tasks Entered Together			
	Intercept		Slope		Intercept		Slope	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Information content	0.004 (0.007)	0.035 (0.009)	0.059 (0.009)	0.031 (0.013)	0.005 (0.011)	0.030 (0.011)	0.020 (0.012)	-0.008 (0.013)
Automation /routine	-0.023 (0.009)	-0.034 (0.008)	-0.070 (0.014)	-0.045 (0.010)	-0.016 (0.011)	-0.030 (0.010)	-0.028 (0.013)	-0.013 (0.012)
No on-site work	0.012 (0.010)	0.015 (0.004)	0.034 (0.007)	0.026 (0.004)	0.003 (0.006)	0.002 (0.005)	0.019 (0.007)	0.021 (0.006)
No face-to-face	-0.030 (0.010)	-0.035 (0.009)	-0.067 (0.018)	-0.051 (0.011)	-0.036 (0.015)	0.002 (0.015)	0.027 (0.017)	-0.014 (0.018)
No decision making	-0.003 (0.009)	-0.035 (0.012)	-0.078 (0.012)	-0.051 (0.015)	0.032 (0.017)	0.001 (0.015)	-0.045 (0.019)	-0.012 (0.018)
Base wage	No	Yes	No	Yes	No	Yes	No	Yes
Adj. R-square	---	---	---	---	0.27	0.51	0.73	0.81

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.

Table 3. Aggregate Decomposition Results

Inequality Measure:	90-10	50-10	90-50	Variance	Gini
A: 1976/78 to 1988/90					
Total Change	0.1742 (0.0077)	0.0700 (0.007)	0.1043 (0.0046)	0.0697 (0.002)	0.0310 (0.0006)
Wage Structure	0.1215 (0.0078)	0.0620 (0.006)	0.0595 (0.0057)	0.0441 (0.0022)	0.0248 (0.0006)
Composition	0.0565 (0.003)	0.0092 (0.0018)	0.0474 (0.0025)	0.0239 (0.001)	0.0054 (0.0003)
Specification Error	-0.0007 (0.006)	0.0011 (0.0051)	-0.0018 (0.0039)	0.0027 (0.0008)	0.0009 (0.0002)
B: 1988/90 to 2000/02					
Total Change	0.0349 (0.0053)	-0.0561 (0.0042)	0.0911 (0.0046)	0.0164 (0.0019)	-0.0023 (0.0005)
Wage Structure	-0.0232 (0.0047)	-0.0802 (0.0039)	0.0570 (0.0044)	-0.0122 (0.0019)	-0.0050 (0.0005)
Composition	0.0443 (0.0017)	0.0195 (0.0012)	0.0248 (0.0016)	0.0203 (0.0006)	0.0005 (0.0003)
Specification Error	0.0156 (0.0024)	0.0059 (0.0022)	0.0097 (0.0018)	0.0091 (0.0005)	0.0023 (0.0001)
C: 2003/04 to 2009/10					
Total Change	0.0480 (0.0065)	0.0224 (0.0041)	0.0256 (0.0063)	0.0156 (0.0024)	0.0042 (0.0006)
Wage Structure	0.0519 (0.0072)	0.0342 (0.0061)	0.0177 (0.0078)	0.0121 (0.0022)	0.0047 (0.0006)
Composition	0.0117 (0.0023)	0.0066 (0.0016)	0.0050 (0.0017)	0.0051 (0.0009)	-0.0001 (0.0003)
Specification Error	-0.0153 (0.0038)	-0.0182 (0.0041)	0.0029 (0.0034)	-0.0016 (0.0006)	-0.0004 (0.0002)

Note: Bootstrapped standard errors are in parentheses (100 replications of the entire procedure). The formulas for the different components are the following and the difference between the total change and the sum of the three components shown is the reweighting error (not shown).

$$\text{Total Change: } \hat{\Delta}_o^v = \overline{\text{RIF}(Y_1, v)} - \overline{\text{RIF}(Y_0, v)} \quad \text{Wage Structure: } \hat{\Delta}_{s,p}^v = \bar{X}_1 (\hat{\gamma}_1^v - \hat{\gamma}_{01}^v)$$

$$\text{Composition: } \hat{\Delta}_{x,p}^v = (\bar{X}_{01} - \bar{X}_0) \hat{\gamma}_0^v \quad \text{Specification Error: } \hat{\Delta}_{x,e}^v = \bar{X}_{01} (\hat{\gamma}_{01}^v - \hat{\gamma}_0^v)$$

Table 4. Detailed Decomposition Results 1988/90-2000/02

Inequality Measure:	90-10	50-10	90-50	Variance	Gini
A: Detailed Composition Effects:					
Union	0.0140 (0.0005)	-0.0132 (0.0004)	0.0272 (0.0008)	0.0062 (0.0002)	0.0034 (0.0001)
Education	-0.0012 (0.0013)	0.0077 (0.0009)	-0.0089 (0.0009)	0.0009 (0.0004)	-0.0034 (0.0001)
Experience	0.0154 (0.0008)	0.0120 (0.0006)	0.0034 (0.0007)	0.0071 (0.0003)	-0.0003 (0.0001)
Technology	0.0022 (0.0006)	0.0070 (0.0005)	-0.0048 (0.0005)	-0.0004 (0.0002)	-0.0009 (0.0001)
Offshorability	0.0109 (0.0007)	0.0063 (0.0006)	0.0046 (0.0005)	0.0053 (0.0003)	0.0008 (0.0001)
Other	0.0030 (0.0003)	-0.0003 (0.0002)	0.0033 (0.0003)	0.0014 (0.0001)	0.0009 (0.0001)
Total	0.0443 (0.0017)	0.0195 (0.0012)	0.0248 (0.0016)	0.0203 (0.0006)	0.0005 (0.0003)
B: Detailed Wage Structure Effects					
Union	0.0051 (0.0017)	0.0018 (0.0013)	0.0033 (0.0016)	0.0029 (0.0005)	0.0020 (0.0001)
Education	0.0621 (0.0076)	0.0057 (0.0063)	0.0563 (0.0073)	0.0253 (0.0028)	0.0049 (0.0007)
Experience	-0.0183 (0.0124)	-0.0045 (0.008)	-0.0144 (0.0119)	-0.0060 (0.005)	-0.0040 (0.0012)
Technology	0.0331 (0.0045)	0.0007 (0.004)	0.0326 (0.0037)	0.0146 (0.0017)	0.0049 (0.0005)
Offshorability	0.0115 (0.0048)	-0.0111 (0.003)	0.0226 (0.0041)	0.0030 (0.0017)	0.0017 (0.0005)
Other	-0.0272 (0.0043)	-0.0214 (0.0037)	-0.0058 (0.0042)	-0.0105 (0.0018)	-0.0037 (0.0005)
Constant	-0.0895 (0.0167)	-0.0514 (0.0135)	-0.0377 (0.0158)	-0.0416 (0.0066)	-0.0109 (0.0017)
Total	-0.0232 (0.0047)	-0.0802 (0.0039)	0.0570 (0.0044)	-0.0122 (0.0019)	-0.0050 (0.0005)

Note: Bootstrapped standard errors are in parentheses (100 replications of the entire procedure).

Figure 1. Changes in Real Wages (\$1979) by Percentile, Men

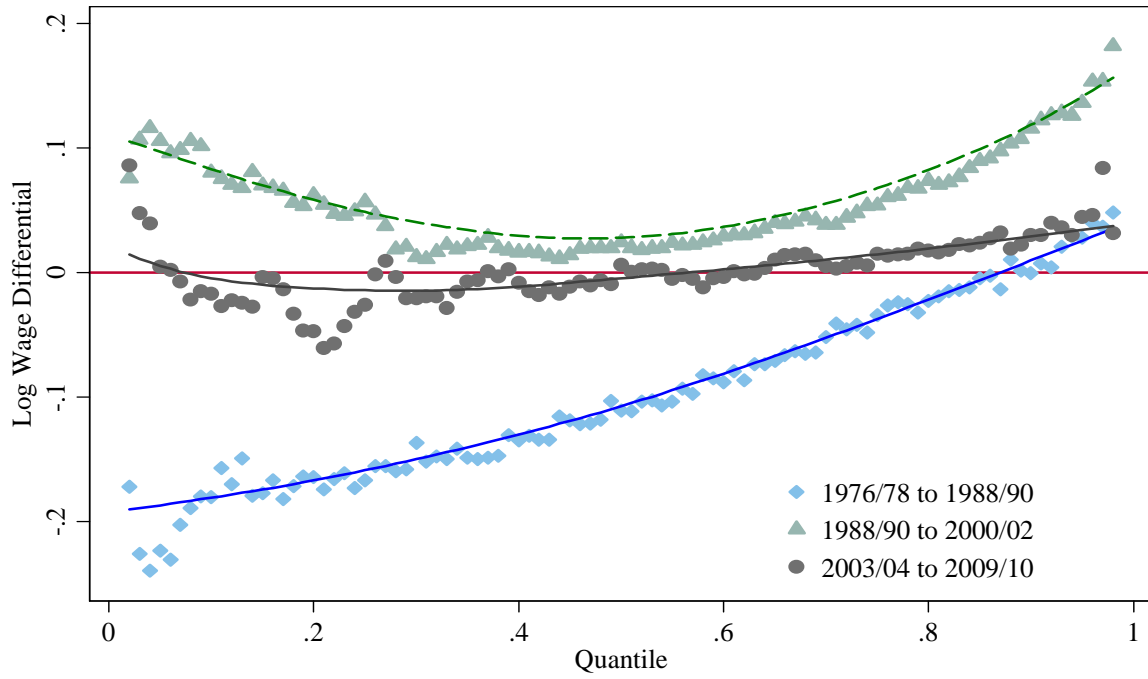


Figure 2. Fitted Changes in Wages from 1988/90 to 2000/02 in Top 30 2-Digit Occupations

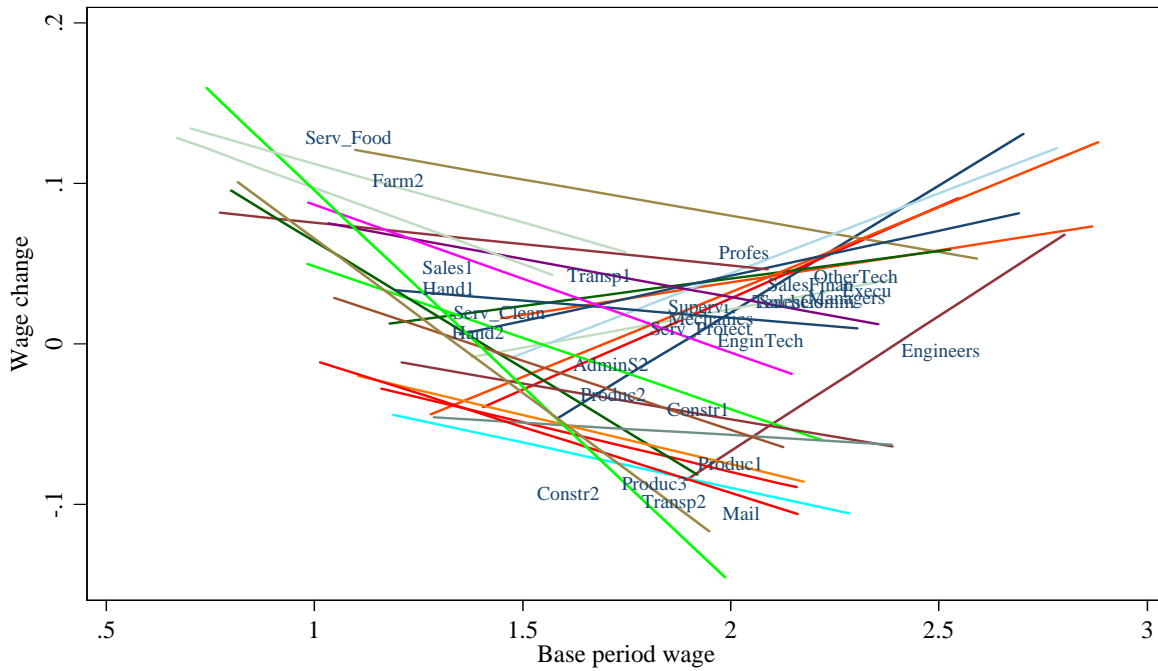




Figure 3. Decomposition of Total Change into Composition and Wage Structure Effects

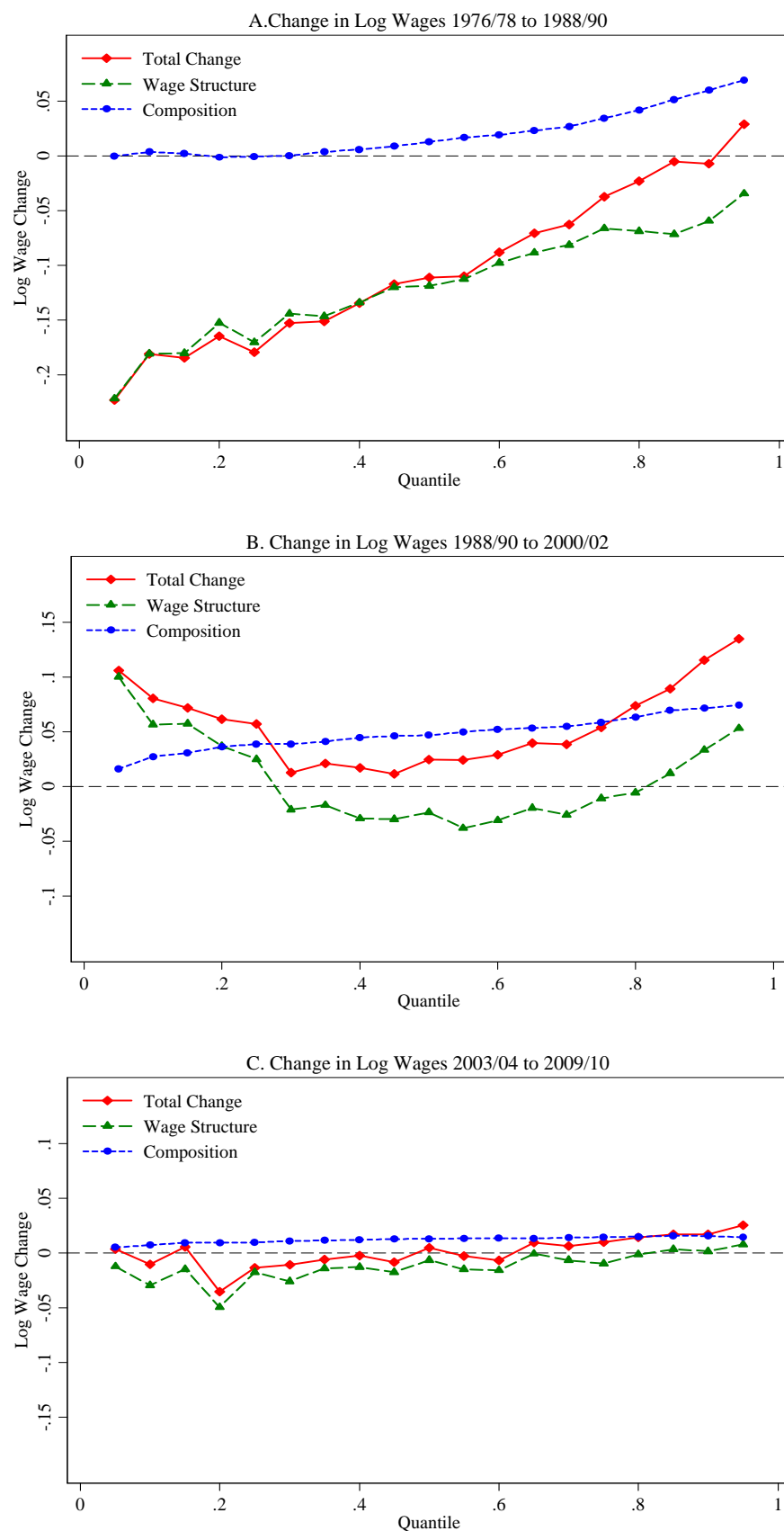


Figure 4. Detailed Decomposition of Composition Effects

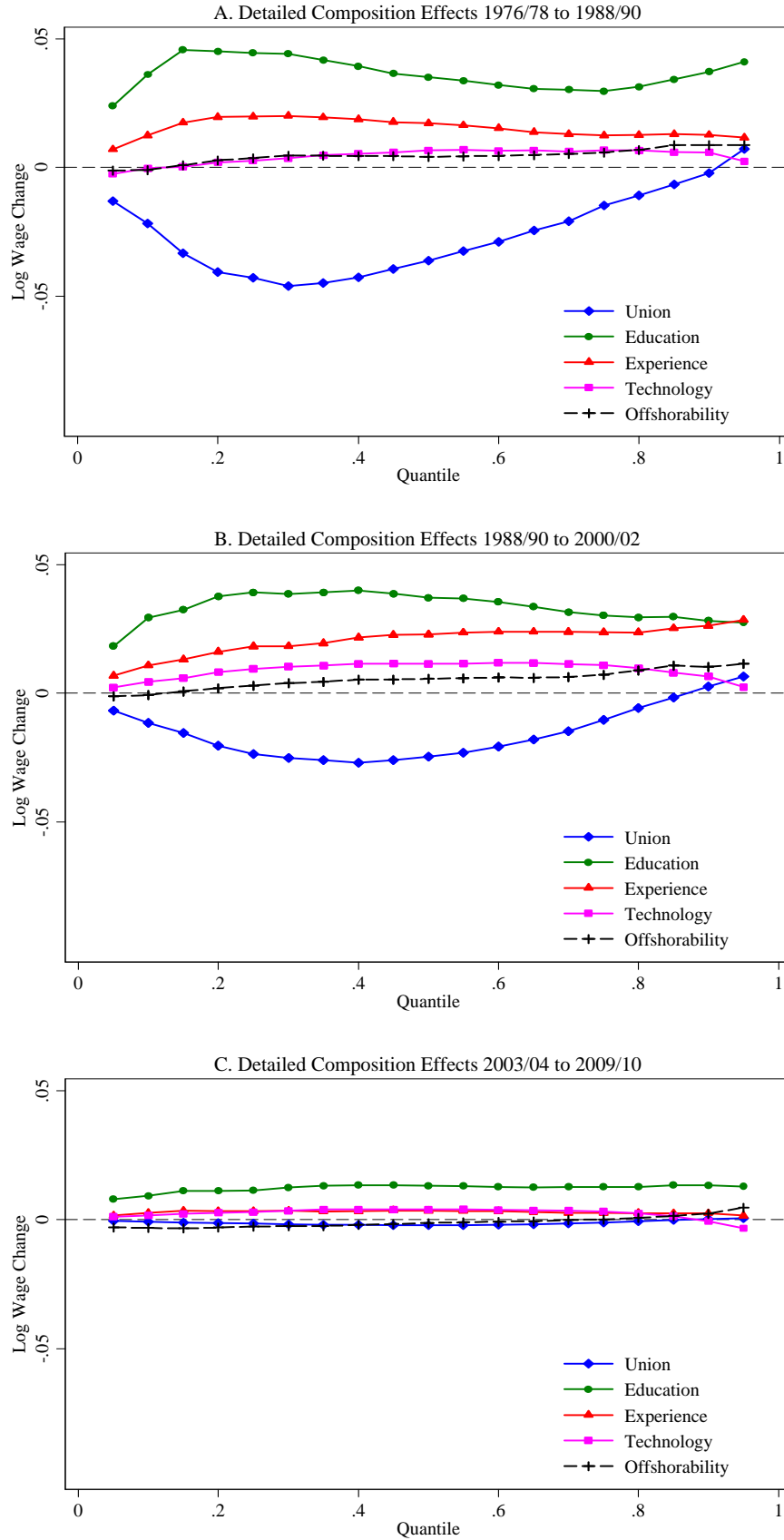
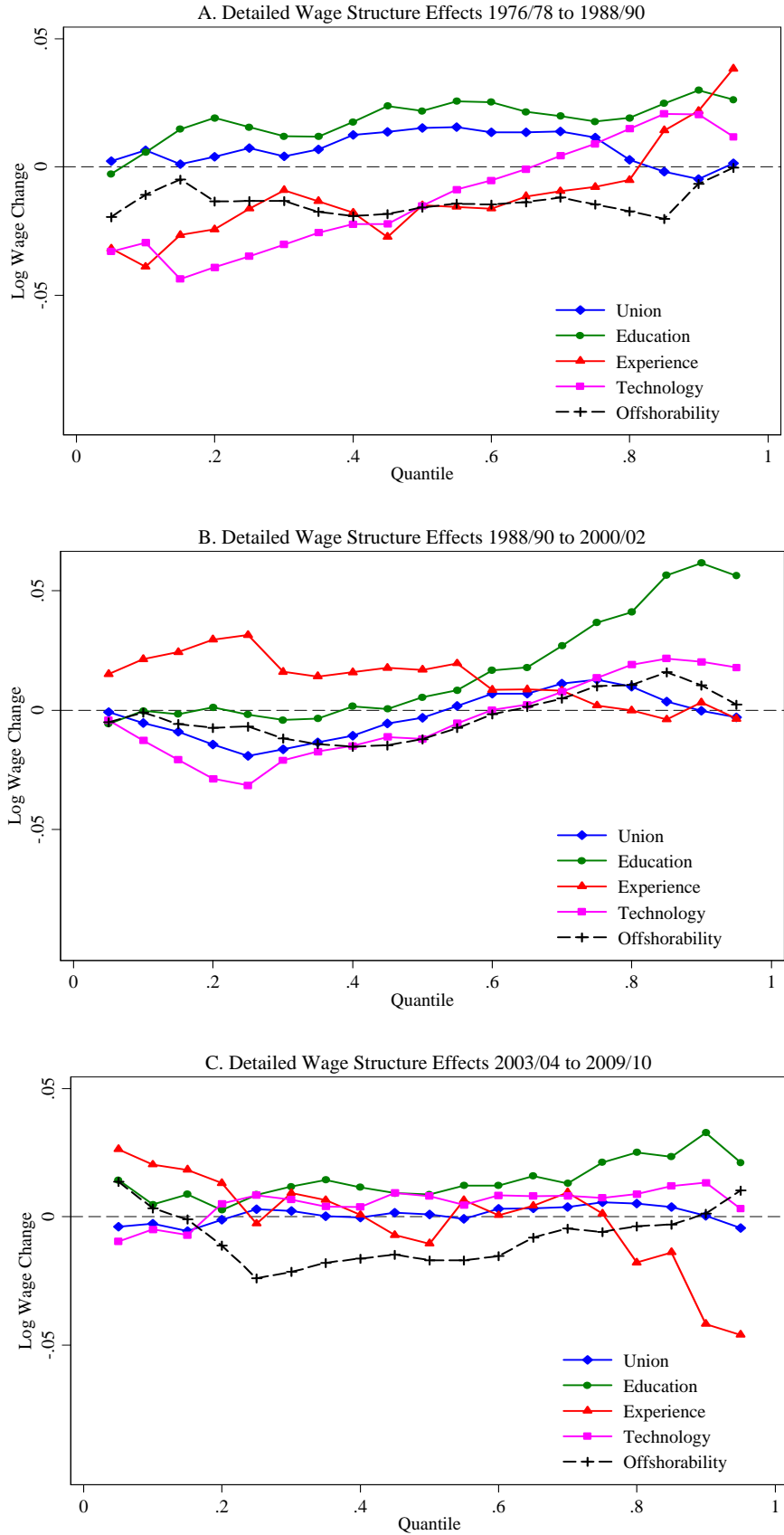


Figure 5. Detailed Decomposition of Wage Structure Effects



**Empirical test of the occupational wage setting model - Technical Appendix**

*A. Wage setting model*

We now explain the empirical approach used to test the two main predictions of the model.

It is useful to first consider how our 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,  $\Delta r_{jk}$ :

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

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

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

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

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

where  $\sigma_{kj}^2$  is the variance of the skill component  $S_{ik}$  for workers in occupation  $k$ , and where  $\sigma_{ujt}^2$  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 (A-1) (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 ( $\Delta r_{jk}$ ). For example, in manual occupations where the return to manual dexterity used to be large ( $r_{jk0} \gg 0$ ) but declined substantially ( $\Delta r_{jk} \ll 0$ ), we expect the slope coefficient  $\tilde{b}_j$  to be negative. By contrast, in some scientific or professional occupations where the return to cognitive skills is high ( $r_{jk0} \gg 0$ ) and does increase over time ( $\Delta r_{jk} \gg 0$ ), we expect the slope coefficient  $\tilde{b}_j$  to be positive.

In our empirical analysis, we rely on large repeated cross-sections of the CPS instead of panel data. While it is not feasible to directly estimate  $\tilde{b}_j$  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  $k = 1, \dots, K$ ), 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

$$(A-2) \quad w_{jt}^q = \bar{w}_{jt} + \sigma_{jt} \Phi^{-1}(q),$$

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

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

Now consider a regression of  $\Delta w_j^q$  on  $w_{j0}^q$ .<sup>43</sup>

$$(A-3) \quad \Delta w_j^q = a_j + b_j w_{j0}^q + e_j^q.$$

The slope parameter,  $b_j$ , is now given by

$$(A-4) \quad b_j = \frac{\text{cov}(\Delta w_j^q, w_{j0}^q)}{\text{var}(w_{j0}^q)} = \frac{(\Delta \sigma_j \cdot \sigma_{j0}) \cdot \text{var}(\Phi^{-1}(q))}{\sigma_{j0}^2 \cdot \text{var}(\Phi^{-1}(q))} = \frac{\Delta \sigma_j}{\sigma_{j0}}.$$

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

$$(A-5) \quad b_j \approx \frac{\sum_{k=1}^K (r_{jk0} \Delta r_{jk}) \cdot \sigma_{kj}^2}{\sigma_{j0}^2} + \frac{\Delta \sigma_{u_j}^2}{2\sigma_{j0}^2}.$$

The first term in equation (A-5) is similar to the slope coefficient obtained earlier in (A-1) 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

<sup>43</sup>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.

relationship between changes in wages and base wage levels.

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

$$(A-6) \quad a_j = \Delta \bar{w}_j - \frac{\Delta \sigma_j}{\sigma_{j0}} \bar{w}_{j0},$$

where

$$(A-7) \quad \Delta \bar{w}_j = \Delta \theta_j + \sum_{k=1}^K \Delta r_{jk} \bar{S}_{jk}.$$

Without loss of generality, we can normalize the base period wage in each occupation to have a mean zero. The intercept can then be written as:

$$(A-8) \quad a_j = \Delta \theta_j + \sum_{k=1}^K \Delta r_{jk} \bar{S}_{jk}.$$

Like the slope parameter  $b_j$ , the intercept  $a_j$  depends on changes in the return to skill components,  $\Delta r_{jk}$ . 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 (A-3) 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 (A-3) 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 ( $\bar{w}_{jt}$  above) and scale parameter ( $\sigma_{jt}$  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 (A-2), this means we can replace  $\Phi^{-1}(q)$  by a more general and occupation-specific inverse probability function  $F_j^{-1}(q)$ . Equation (A-2) is then replaced by

$$(A-9) \quad w_{jt}^q = \bar{w}_{jt} + \sigma_{jt} F_j^{-1}(q).$$

We can then get the same regression equation since

$$(A-10) \quad \Delta w_j^q = \Delta \bar{w}_j + \Delta \sigma_j F_j^{-1}(q).$$

Solving for  $F_j^{-1}(q)$  in equation (A-9) at  $t = 0$ , and substituting into equation (A-10) yields

$$(A-11) \quad \Delta w_j^q = \Delta \bar{w}_j - \frac{\Delta \sigma_j}{\sigma_{j0}} \bar{w}_{j0} + \frac{\Delta \sigma_j}{\sigma_{j0}} w_{j0}^q,$$

which is identical to equation (A-3) since  $a_j = \Delta \bar{w}_j - \frac{\Delta \sigma_j}{\sigma_{j0}} \bar{w}_{j0}$  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 (A-10) becomes

$$(A-12) \quad \Delta w_j^q = \Delta \bar{w}_j + \left( \sigma_{j1} F_{j1}^{-1}(q) - \sigma_{j0} F_{j0}^{-1}(q) \right)$$

$$(A-13) \quad = \Delta \bar{w}_j + \Delta \sigma_j F_{j0}^{-1}(q) + e_j^q$$

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

$$(A-14) \quad \Delta w_j^q = a_j + b_j w_{j0}^q + e_j^q.$$

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  $\lambda^q$  in  $e_j^q$ :

$$(A-15) \quad e_j^q = \lambda^q + \varepsilon_j^q.$$

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

$$(A-16) \quad \Delta w_j^q = a_j + b_j w_{j0}^q + \lambda^q + \varepsilon_j^q.$$

A more economically intuitive interpretation of the percentile-specific error components  $\lambda^q$  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  $\lambda^q$ .

### *B. Estimation results*

The estimates of equation (3) are reported in Appendix Table A3. We use two different measures of fit to evaluate how the model performs empirically. The adjusted R-square of the model is first used to assess whether model in equation (3) is the true wage determination model. We note that because of the residual sampling error in the estimated wage changes, the regressions would not explain all of the variation in the data.<sup>44</sup> 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 1. For each estimated model, we run a simple regression of the estimated residuals on a linear and quadratic term in the base wage

$$(A-17) \quad \widehat{\varepsilon}_j^q = \pi_0 + \pi_1 w_{j0}^q + \pi_2 (w_{j0}^q)^2 + v_j^q,$$

to see whether there is any curvature left in the residuals that the model is unable to explain. As a reference, the coefficient on the quadratic term,  $\widehat{\pi}_2$ , in the fitted regression on the raw data is equal to 0.136.

We find that when equation (3) is estimated with the base wage as only explanatory variable (column 1), essentially none of the variation in the data is explained as the adjusted R-square is only equal to 0.0218. Since a linear regression cannot explain any of the curvature of the changes in wages, the curvature parameter in the residuals (0.136) is exactly the same as in the simple

<sup>44</sup>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.



quadratic regression discussed above. When we only include the set of occupation dummies (the  $a_j$ 's) in the regression in column 2, we explain more than half of the raw variation in the data, and about a third of the curvature. The curvature parameter declines from 0.136 to 0.087 but remains strongly significant. 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.

When we include only the decile dummies (the  $\lambda^q$ 's) in column 3, essentially none of the variation or curvature in the data is explained. 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. 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. 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 (the  $b_j$ 's) in column 5. The R-square of the model jumps to 0.9274, 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  $\hat{\pi}_2$  now drops to 0.002 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 dummies play a more substantial role as we see a drop in the adjusted R-square from 0.9274 (decile dummies included) to 0.8602 (decile dummies excluded).

Finally, Panel B of Appendix Table A3 shows the results of similar specifications using a reweighting approach to control for standard covariates. In that case, we also find that the model with decile dummies and occupation-specific slopes explains most of the variation in the data and all of the curvature, although the R-square is generally lower since the covariates reduce the explanatory power of occupations.

Appendix Table A1. Descriptive Statistics

	1976/78		1988/90		Difference in Means (88/90- 76/78)	2000/02		Difference in Means (00/02-88/90)
	Means	Standard Deviation	Means	Standard Deviation		Means	Standard Deviation	
Log wages	1.851	0.520	1.753	0.583	-0.098	1.812	0.597	0.059
Union covered	0.295	0.456	0.202	0.401	-0.093	0.149	0.356	-0.053
Non-white	0.101	0.302	0.127	0.333	0.026	0.140	0.347	0.013
Non-Married	0.295	0.456	0.386	0.487	0.091	0.415	0.493	0.028
Education								
Primary	0.103	0.303	0.060	0.237	-0.043	0.042	0.200	-0.018
Some HS	0.174	0.379	0.121	0.326	-0.053	0.089	0.285	-0.032
High School	0.369	0.483	0.379	0.485	0.009	0.312	0.463	-0.067
Some College	0.168	0.374	0.203	0.402	0.035	0.274	0.446	0.071
College	0.106	0.307	0.137	0.344	0.032	0.188	0.391	0.051
Post-grad	0.080	0.272	0.100	0.301	0.020	0.095	0.294	-0.005
Age	35.708	12.854	35.766	11.738	0.058	37.569	11.824	1.803

	2003/04		2009/10		Difference in Means
	Means	Standard Deviation	Means	Standard Deviation	
Log wages	1.819	0.606	1.820	0.616	0.001
Union covered	0.141	0.348	0.136	0.343	-0.005
Non-white	0.133	0.340	0.144	0.351	0.011
Non-Married	0.406	0.491	0.423	0.494	0.017
Education					
Primary	0.045	0.207	0.039	0.193	-0.006
Some HS	0.083	0.277	0.071	0.257	-0.013
High School	0.309	0.462	0.302	0.459	-0.007
Some College	0.270	0.444	0.276	0.447	0.006
College	0.194	0.396	0.207	0.405	0.012
Post-grad	0.099	0.298	0.106	0.308	0.008
Age	38.166	11.956	39.150	12.299	0.984

Appendix Table A2. O\*NET 13.0 — Work Activities & Work Context

A. Characteristics linked to Technological Change/Offshorability<sup>1</sup>

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.2.1 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

---

<sup>1</sup> Note: (JK) indicates a work activity used in Jensen and Kletzer (2007), (B) a work activity used or suggested in Blinder (2007).

Appendix Table A3: Regression Fit of Models for 1988/90 to 2000/02 Changes in Wages at each Decile, by 2-Digit Occupation

	(1)	(2)	(3)	(4)	(5)	(6)
<b>A. Models without controls for observables</b>						
Adj. R-square	0.0218	0.5535	0.0284	0.5996	0.9274	0.8602
Curvature in residuals ( $\beta_2$ )	0.136 (0.009)	0.087 (0.005)	0.122 (0.009)	0.073 (0.006)	0.002 (0.003)	0.020 (0.004)
<b>B. Models with controls for observables</b>						
Adj. R-square	0.0730	0.3982	0.0498	0.4644	0.8743	0.7711
Curvature in residuals ( $\beta_2$ )	0.131 (0.015)	0.116 (0.011)	0.084 (0.015)	0.068 (0.011)	0.002 (0.005)	0.052 (0.007)
Occupation dummies (Intercept)		Yes		Yes	Yes	Yes
Decile dummies			Yes	Yes	Yes	
Base wage	Yes				Yes	Yes
Occ * base wage (Slope)					Yes	Yes

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 observable. 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.

Appendix Table A4. Unconditional Quantile Regression Coefficients on Log Wages

Explanatory Variables	Years:	1988/90			2000/02		
	Quantiles:	10	50	90	10	50	90
Union covered		0.218 (0.003)	0.454 (0.005)	-0.048 (0.005)	0.161 (0.004)	0.414 (0.006)	-0.091 (0.007)
Non-white		-0.070 (0.008)	-0.136 (0.005)	-0.080 (0.005)	-0.037 (0.006)	-0.126 (0.005)	-0.045 (0.008)
Non-Married		-0.152 (0.005)	-0.127 (0.004)	-0.036 (0.005)	-0.095 (0.004)	-0.142 (0.005)	-0.089 (0.006)
Education ( Some College omitted)							
Primary		-0.443 (0.014)	-0.504 (0.008)	-0.220 (0.006)	-0.496 (0.018)	-0.519 (0.01)	-0.134 (0.007)
Some HS		-0.431 (0.011)	-0.271 (0.006)	-0.089 (0.005)	-0.443 (0.01)	-0.300 (0.006)	-0.015 (0.005)
High School		-0.051 (0.006)	-0.134 (0.005)	-0.106 (0.005)	-0.047 (0.005)	-0.157 (0.005)	-0.072 (0.006)
College		0.103 (0.005)	0.220 (0.006)	0.338 (0.009)	0.063 (0.005)	0.248 (0.006)	0.449 (0.01)
Post-grad		0.042 (0.005)	0.230 (0.006)	0.665 (0.014)	0.022 (0.006)	0.278 (0.007)	1.006 (0.022)
Potential Experience (15< Experience < 20 omitted)							
Experience <5		-0.559 (0.009)	-0.472 (0.007)	-0.337 (0.009)	-0.438 (0.009)	-0.414 (0.008)	-0.247 (0.01)
5< Experience < 10		-0.067 (0.007)	-0.283 (0.006)	-0.303 (0.009)	-0.062 (0.007)	-0.259 (0.007)	-0.278 (0.011)
10< Experience < 15		-0.016 (0.007)	-0.127 (0.006)	-0.190 (0.01)	-0.027 (0.007)	-0.108 (0.007)	-0.140 (0.013)
20< Experience < 25		-0.002 (0.006)	-0.054 (0.006)	-0.102 (0.01)	-0.012 (0.006)	-0.049 (0.007)	-0.030 (0.011)
25< Experience < 30		(0.01) (0.006)	(0.032) (0.007)	(0.06) (0.012)	-(0.002) (0.006)	(0.023) (0.007)	(0.007) (0.012)
30< Experience < 35		0.017 (0.008)	0.045 (0.008)	0.060 (0.011)	-0.003 (0.006)	0.023 (0.007)	0.019 (0.014)
35< Experience < 40		0.023 (0.009)	0.021 (0.008)	0.048 (0.012)	0.004 (0.007)	0.008 (0.009)	0.035 (0.015)
Experience > 40		0.085 (0.009)	0.015 (0.008)	-0.027 (0.01)	-0.007 (0.012)	-0.044 (0.01)	-0.042 (0.014)
O*NET Measures							
Information Content		0.067 (0.003)	0.086 (0.002)	0.023 (0.004)	0.052 (0.003)	0.096 (0.003)	0.044 (0.004)
Automation		0.015 (0.002)	-0.035 (0.002)	-0.044 (0.002)	0.014 (0.002)	-0.055 (0.002)	-0.023 (0.003)
No Face-to-Face		0.114 (0.003)	0.122 (0.002)	0.115 (0.003)	0.086 (0.002)	0.120 (0.002)	0.121 (0.005)
Non On-Site Job		-0.027 (0.002)	0.050 (0.002)	0.092 (0.003)	-0.028 (0.002)	0.044 (0.002)	0.104 (0.003)
No Decision-Making		-0.157 (0.004)	-0.148 (0.003)	-0.142 (0.004)	-0.136 (0.003)	-0.157 (0.004)	-0.137 (0.005)
Constant		1.598 (0.011)	2.774 (0.012)	2.465 (0.015)	1.219 (0.006)	1.896 (0.006)	2.524 (0.01)
Number of obs.			226,076			167,929	

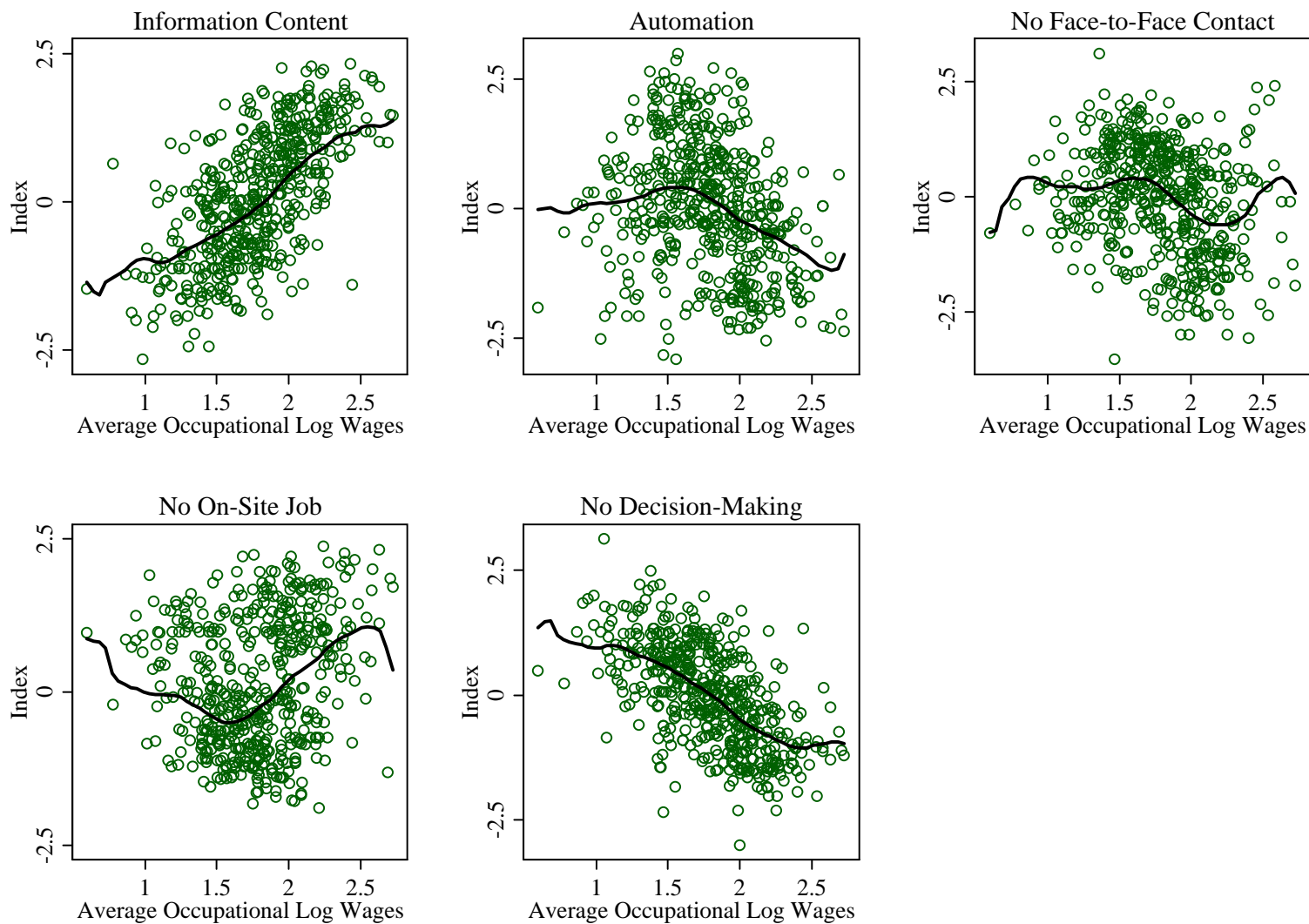
Note: Bootstrapped standard errors are in parentheses (100 replications of the entire procedure).

Appendix Table A5. RIF Regression of Inequality Measures on Log Wages

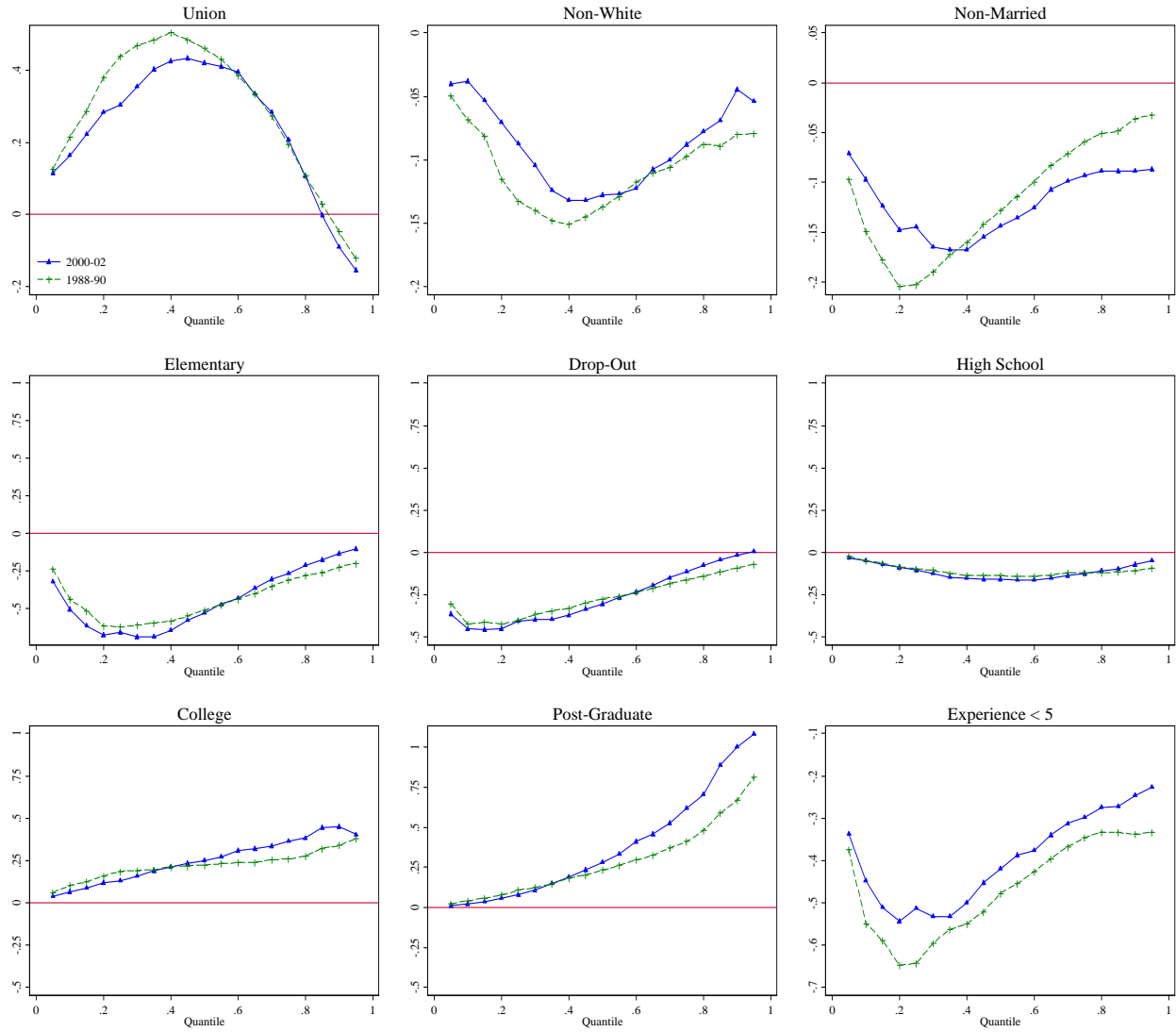
Dependent Variables	Variance:	0.3401	0.3565	Gini:	0.1883	0.1860
Explanatory Variables	Years:	1988/90	2000/02	1988/90	2000/02	
Union covered		-0.1154 (0.0023)	-0.1133 (0.0026)	-0.0630 (0.0007)	-0.0506 (0.0007)	
Non-white		-0.0057 (0.0027)	-0.0059 (0.004)	0.0106 (0.0009)	0.0072 (0.001)	
Non-Married		0.0510 (0.0022)	0.0157 (0.0032)	0.0277 (0.0007)	0.0163 (0.0006)	
Education (Some College omitted)						
Primary		0.0829 (0.0046)	0.1595 (0.0069)	0.0717 (0.0018)	0.0883 (0.0021)	
Some HS		0.1037 (0.0035)	0.1548 (0.0041)	0.0594 (0.0012)	0.0697 (0.0014)	
High School		-0.0271 (0.0025)	-0.0056 (0.0029)	0.0038 (0.0009)	0.0085 (0.0007)	
College		0.0977 (0.0046)	0.1458 (0.0049)	0.0031 (0.0012)	0.0153 (0.001)	
Post-grad		0.2670 (0.0052)	0.3903 (0.0075)	0.0433 (0.0012)	0.0659 (0.0015)	
Potential Experience (15 < Experience < 20 omitted)						
Experience < 5		0.0636 (0.0043)	0.0803 (0.0056)	0.0709 (0.0013)	0.0623 (0.0014)	
5 < Experience < 10		-0.0993 (0.0041)	-0.0801 (0.0045)	-0.0028 (0.0012)	-0.0012 (0.0012)	
10 < Experience < 15		-0.0782 (0.0042)	-0.0466 (0.0048)	-0.0099 (0.0011)	-0.0035 (0.0011)	
20 < Experience < 25		-0.0522 (0.0045)	-0.0107 (0.0055)	-0.0086 (0.0011)	0.0008 (0.0013)	
25 < Experience < 30		0.0222 (0.0051)	0.0021 (0.005)	0.0028 (0.0013)	-0.0003 (0.0013)	
30 < Experience < 35		0.0167 (0.0061)	0.0086 (0.0053)	-0.0001 (0.0015)	0.0014 (0.0013)	
35 < Experience < 40		0.0115 (0.006)	0.0217 (0.0068)	0.0002 (0.0015)	0.0046 (0.0015)	
Experience > 40		-0.0407 (0.0059)	-0.0010 (0.0071)	-0.0143 (0.0016)	0.0035 (0.002)	
O*NET Measures						
Information Content		-0.0263 (0.0015)	-0.0130 (0.0025)	-0.0142 (0.0005)	-0.0094 (0.0005)	
Automation		-0.0196 (0.0012)	-0.0144 (0.0015)	-0.0037 (0.0004)	-0.0022 (0.0004)	
No Face-to-Face		-0.0079 (0.0018)	-0.0029 (0.0016)	-0.0145 (0.0004)	-0.0106 (0.0004)	
Non On-Site Job		0.0439 (0.0014)	0.0490 (0.0017)	0.0082 (0.0004)	0.0097 (0.0004)	
No Decision-Making		0.0012 (0.0018)	0.0052 (0.0024)	0.0172 (0.0006)	0.0166 (0.0005)	
Constant		0.0986 (0.0059)	0.2705 (0.0046)	0.0445 (0.0019)	0.1528 (0.0012)	

Note: Bootstrapped standard errors are in parentheses (100 replications of the entire procedure).

Appendix Figure A1. Average Occupational Wages in 2002/02 by Task Category Indexes

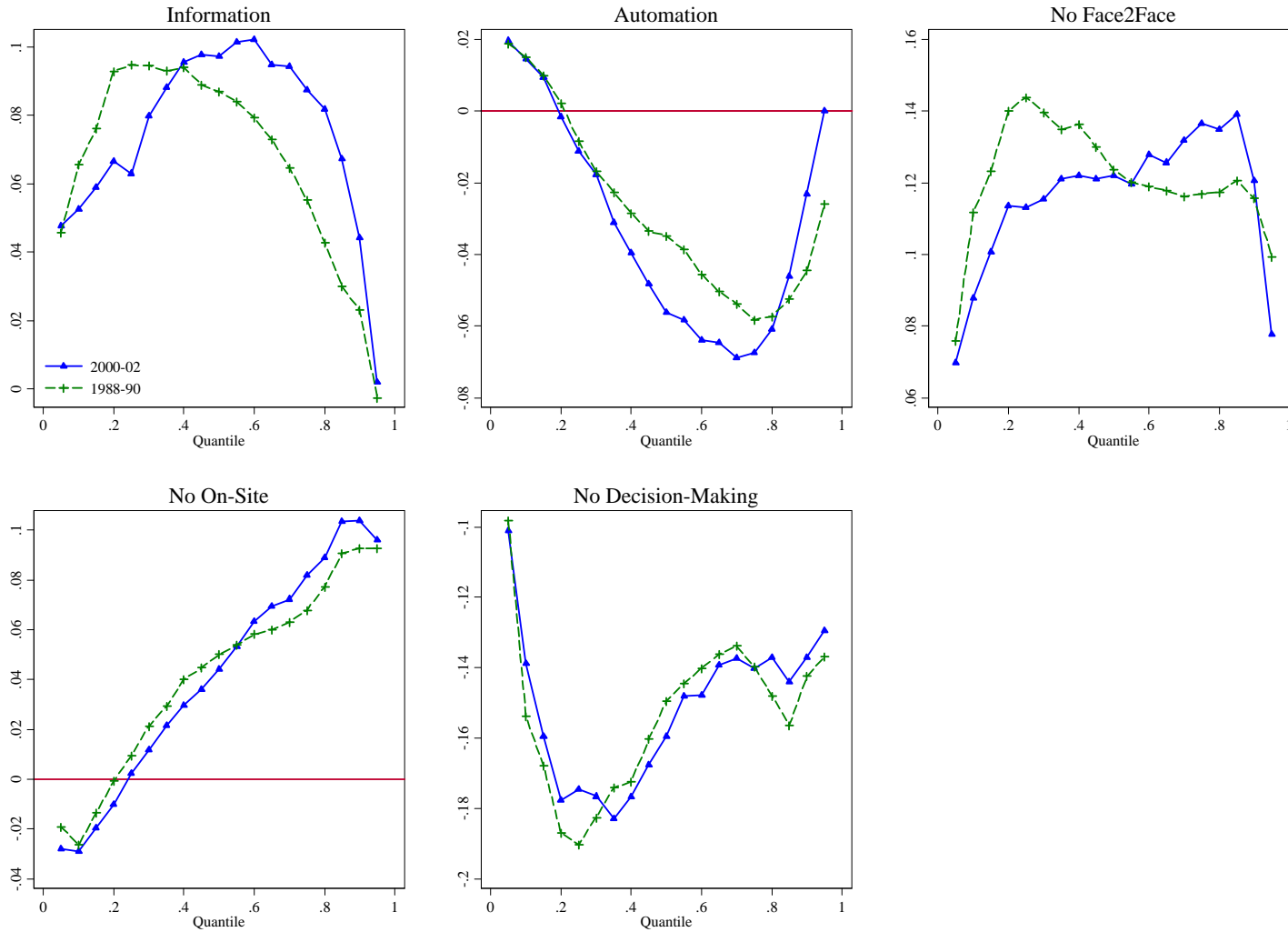


Appendix Figure A2. Unconditional Quantile Regressions Coefficients:  
Selected Demographic Variables 1988/90-2000/02

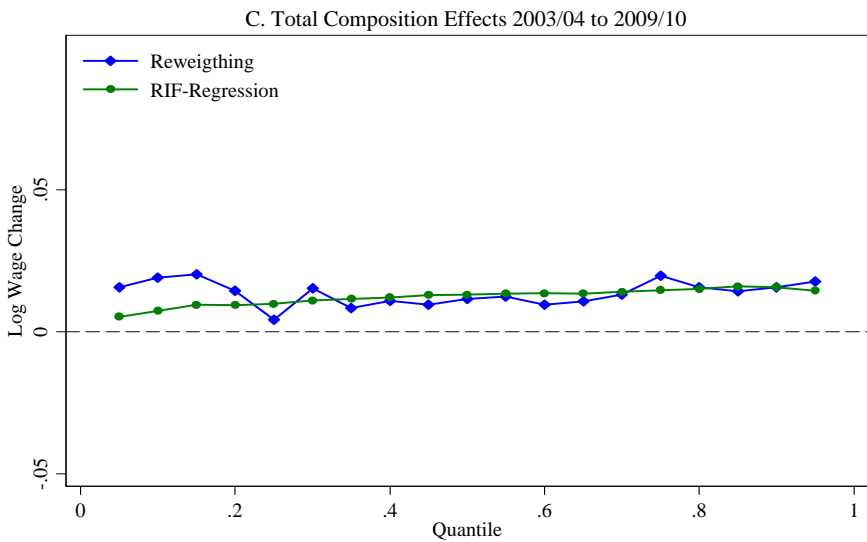
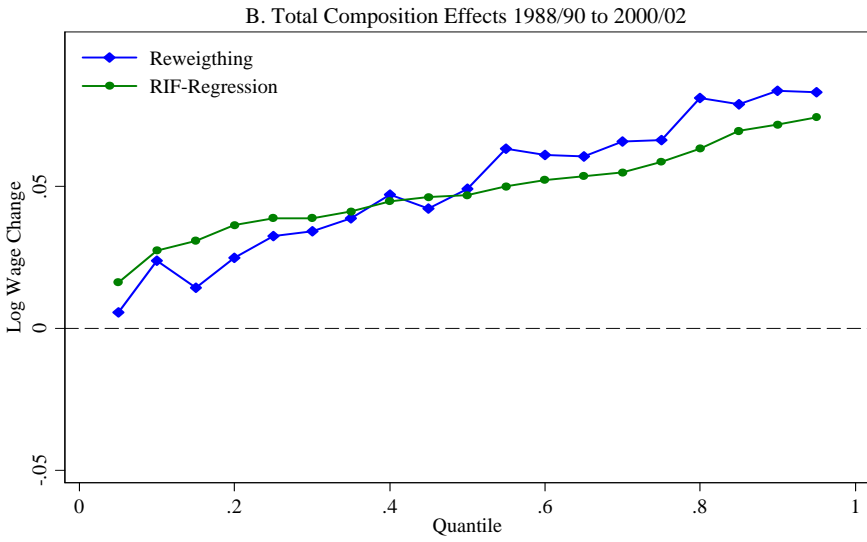
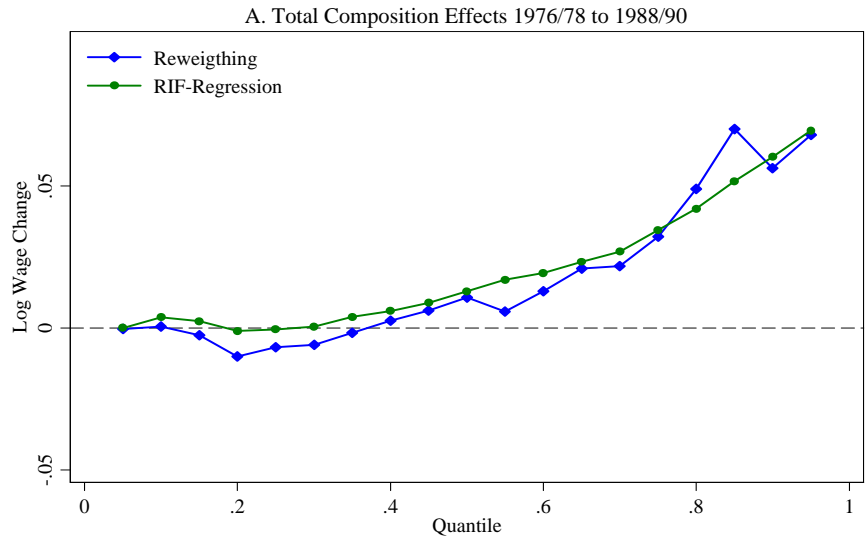




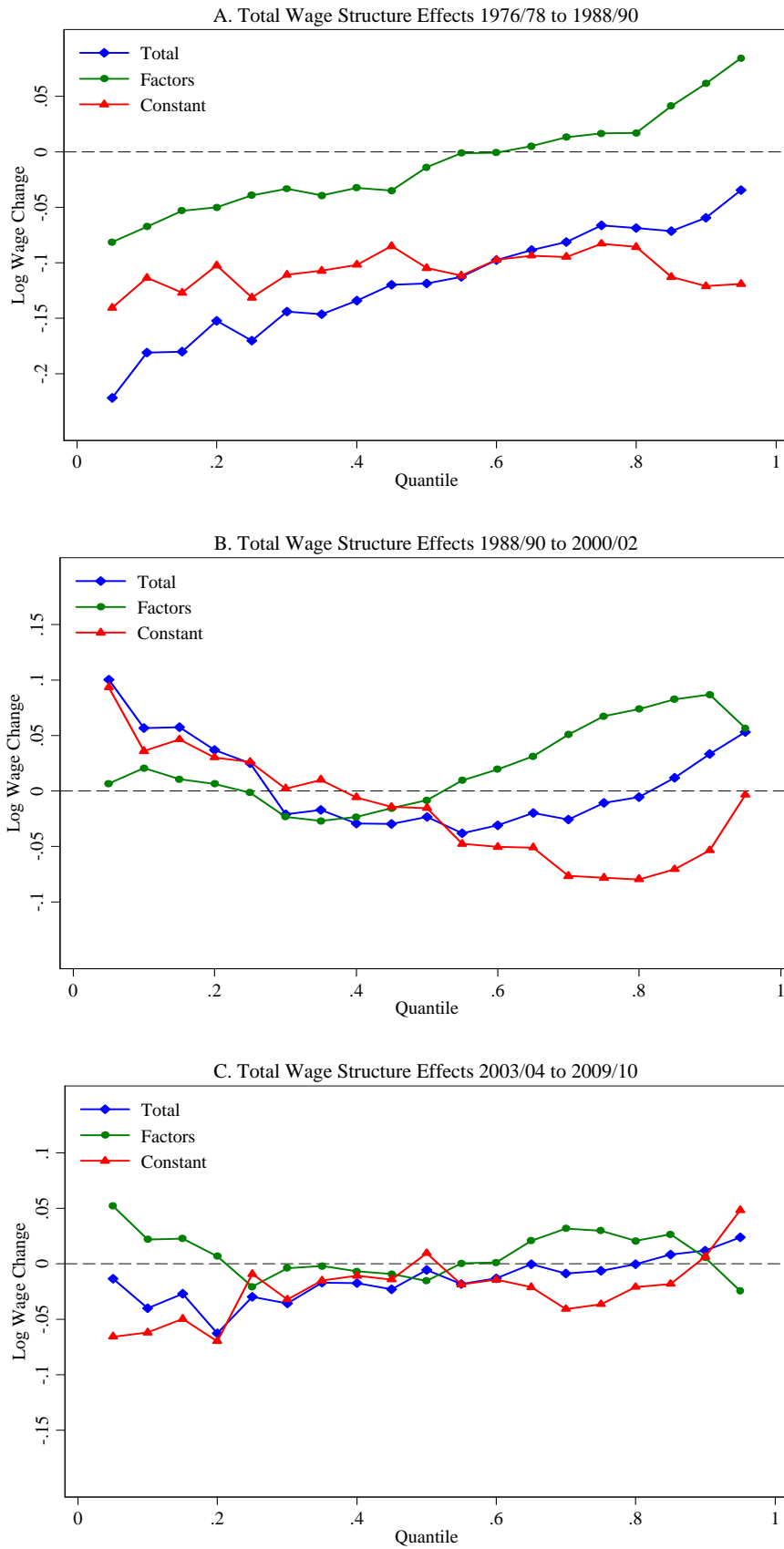
Appendix Figure A3. Unconditional Quantile Regressions Coefficients:  
Occupational Tasks 1988/90-2000/02



# Appendix Figure A4. Total Composition Effects



## Appendix Figure A5. Total Wage Structure Effects



Appendix Figure A6. Detailed Decomposition Effects with Minimum Wages

