Inequality and Changes in Task Prices: Within and Between Occupation Effects

Nicole Fortin

Vancouver School of Economics, UBC and IZA

Thomas Lemieux

Vancouver School of Economics, UBC and IZA, NBER

March, 2015

Abstract

This paper looks at changes in the structure of wages at the occupation level, and connects those to measures of the task content of jobs. We first present a simple model where skills are used to produce tasks, and changes in task prices are the underlying source of change in occupational wages. Using Current Population Survey and task measures from the O*NET, we document large changes in both the between and within dimension of occupational wages over time, and argue that these changes are well explained by changes in task prices induced by offshoring and technological change.

1 Introduction

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

This situation has changed dramatically in recent years. Starting with the highly influential work of Autor, Levy and Murnane (2003), the literature has paid increasingly more attention to the role of tasks and occupations in changes in the wage structure. There is now a growing body of work recently summarized by Acemoglu and Autor (2011) that goes beyond the standard model of skills and wages to formally incorporate the role of tasks and occupations in changes in the wage distribution. Despite this recent work, however, there is still limited work trying to look explicitly at how changes in returns to tasks or occupations have contributed to changes in the overall wage structure. The main contribution of this paper is to help close this gap by documenting how the occupational wage structure have changed over time, and how this is connected to measures of task content in these occupations.

The paper is organized as follows. In Section 2, we present a simple model where returns to a variety of skills are different in different occupations. This model provides a rationale for connecting the task content of occupations with wage setting in these occupations. In Section 3 we introduce measures of task content computed from the O*Net data, and explain how we link those to various sources of change in task prices, such as technological change and offshoring. Section 4 documents changes in the level and dispersion of wages across occupations, and looks at how these changes are connected to our measures of the task content of jobs. We conclude in Section 5.

2 Wage Setting in Occupations

This section relies heavily on Firpo, Fortin, and Lemieux (2013) who use a similar model to perform an exhaustive decomposition of changes in the wage structure between the late 1970s and recent years. They focus on the contribution of occupational tasks, as

measured using the O*NET data, in the overall changes in wage inequality. In Firpo, Fortin, and Lemieux (2013), the key mechanism involved is that changes in task prices affect the whole pricing structure for each occupation, which then contributes to changes in wage inequality. Their decomposition approach allows them to aggregate the impact of all these changes in occupation pricing on the overall wage distribution. In this paper, we instead focus on the implicit first step in this approach, i.e. the effect of changes in task prices on the occupational wage structure.

To fix ideas, it is useful to remember that, until recently, most of the wage inequality literature has followed 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 factors like 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 approach as the "canonical model" that has been used in many influential studies, such as Katz and Murphy (1992).

There is increasing evidence that the canonical model does not provide a satisfactory explanation for several important features of the evolution of 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. First, it cannot account for differential changes in inequality in different parts of the distribution, such as the "polarization" of the wage distribution of the 1990s illustrated in Figure 1. 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 where workers use their skills to produce tasks, and get systematically allocated to occupations (i.e. tasks) on the basis of comparative advantage.[1]

We closely follow Acemoglu and Autor (2011) in the way we introduce the distinction between skills and tasks in our wage setting model. Unlike Acemoglu and Autor (2011), however, we do not attempt to solve the full model of skills, tasks, and wages by modelling how workers choose occupations, and how supply and demand shocks affect wages in general equilibrium. One advantage of our partial equilibrium approach is that we don't

---

[1] Note that since different tasks are being performed in different occupations, we can think of these two concepts interchangeably.

have to impose restrictive assumptions to help solve the model. For instance, Acemoglu and Autor (2011) have to work with only three skill groups (but many occupations/tasks) to get interesting predictions out of their model. As a result, the law of one price holds within each skill group in the sense that wages are equalized across occupations, conditional on skill. This is a strong prediction that is not supported in the data, and that we can relax by allowing for a large number of skill categories.[2] This limits our ability to solve the model in general equilibrium, which is beyond the scope of this paper. Yet, the fact that workers systematically sort into different occupations/tasks on the basis of their skills has potentially important implications for the interpretation of our results. We discuss these issues in more detail at the end of this section.

Like Acemoglu and Autor (2011), we assume that an occupation j involves producing a task or occupation-specific output Y_j which is one input in the firm's production function. But instead of just working with three skill types, we assume that workers are characterized by a k-dimension set of skills $S_i = [S_{i1}, S_{i2}, ..., S_{iK}]$. Some of these skills (like education and experience) are observed by the econometrician, others (like ability and motivation) are not. The amount of occupation-specific task Y_{ij} produced by worker i in occupation j is assumed to linearly depend on skill:

$$Y_{ij} = \sum_{k=1}^{K} \alpha_{jk} S_{ik}, \tag{1}$$

where the productivity of skills α_{jk} are specific to occupation j. Firms then combine tasks to produce final goods and services according to the production function $Q = F(Y_1, ..., Y_J)$ where Y_j (for $j = 1, .., J$) is the total amount of (occupation-specific) tasks produced by all workers i allocated to occupation j.[3]

Under the assumption that wages are set competitively, workers are paid for the value of tasks they produce. Worker i who produces Y_{ij} units of occupation-specific task j is thus paid a wage of $p_{jt} Y_{ij}$, where p_{jt} is the market price of each unit of task Y_{ij} produced at time t. We also allow wages to depend on year and occupation specific factors δ_t and c_j, where δ_t could capture, for instance, general productivity shocks, while c_j could be thought as reflecting compensating wage differentials. In the empirical analysis, we also consider other factors Z_{it} such as institutions (e.g. union status) and discrimination (e.g.

---

[2]See, for instance, Heckman and Scheinkman (1987) and Gibbons et al. (2005) for evidence of occupational wage differences among workers with similar observed and unobserved productive characteristics.

[3]This specification is also closely related to the "skill-weights" approach of Lazear (2009) where different jobs require the use of different linear combinations of skills.

race and gender) that affect wages in a way that is unrelated to task output. This yields the wage equation:

$$w_{ijt} = \delta_t + Z_{it}\psi_t + c_j + p_{jt}Y_{ij} \equiv \delta_t + c_j + Z_{it}\psi_t + p_{jt}\sum_{k=1}^{K} \alpha_{jk}S_{ik}. \qquad (2)$$

As in Acemoglu and Autor (2011), a critical assumption embedded into equation (2) is that the mapping of skills into tasks (the parameters α_{jk} in the wage equation) does not change over time, while task prices p_{jt} are allowed to change over time. This means that, in this model, the effect of demand factors such as offshoring and technological change solely goes through changes in task prices. In this setting, technological change and offshoring provide a way for firms of producing the same tasks at a lower price. Take, for instance, the case of call center operators who use their skills to produce consumer service tasks (check customer accounts, provide information about products, etc.). When these tasks are simple, like providing one's balance on a credit card, the call center operators can be replaced by computers now that voice recognition technology is advanced enough. In the case of more complex tasks such as IT support, computers are not sophisticated enough to deal with customers but these tasks can now be offshored to lower paid workers in India. In these examples, the quantity of task produced by call center operators of a given skill level does not change, but the wage associated with these tasks changes in response to technological change and offshoring. At the limit, if the task price in an occupation becomes low enough the occupation will simply disappear, which is the way Acemoglu and Autor (2011) model the impact of "routine-biased" technological change.

In other cases the assumption that the mapping between skills and tasks is constant over time may be unrealistic. For instance, in highly technical or professional occupations where cognitive skills are important for producing tasks, advances in computing likely enable workers with a given set of skills to produce more tasks than they used to. In that setting, when wages increase for these workers, equation (2) would suggest that task prices have increased, while the underlying explanation may instead be productivity changes linked to changes in the α_{jk}'s. Since p_{jt} and α_{jk} enter multiplicatively in equation (2), it is not possible to empirically distinguish the impact of changes in these two factors. Ultimately, the product of p_{jt} and α_{jk} is an occupation-specific return to skill at time t, and the main goal of the paper is to quantify the contribution of changes in the these occupation-specific returns to skill on changes in the wage distribution, controlling for other factors usually considered in the inequality literature. For the sake of simplicity we interpret these changes in returns as changes in task prices, but acknowledge that they

could also reflect occupation-specific productivity effects.

When task prices are allowed to vary across occupations in a completely unrestricted way, it is difficult to interpret the contribution of changes in task prices to changes in inequality in an economically meaningful way. Following Yamaguchi (2012), we assume that task prices are systematically linked to a limited number of task content measures available in data sets like the Dictionary of Occupational Titles or the O*NET. The idea is that two different occupations where the task content measure for, say, "routine work" is the same will be equally affected by "routine-biased" technological change. In the empirical part of the paper we use a set of five task content measures from the O*NET that are described in detail in the next section. We use the following linear specification for task prices:

$$p_{jt} = \pi_{0t} + \sum_{h=1}^{5} \pi_{ht} T_{jh} + \mu_{jt}, \tag{3}$$

where T_{jh} are the task content measures. These task content measures are assumed to be time invariant for two reasons. First, it has proven difficult to construct consistent measures of the task content of occupations over time because of data limitations (see, e.g., Autor, 2013). More importantly, we use the task content measures as an economically interpretable way of reducing the dimension of the occupational space. Results would be hard to interpret if the way in which task content characterized occupations was also changing over time.[4]

Since the T_{jh}'s do not change over time, changes in task prices p_{jt} are solely due to change in the parameters π in equation (3). These parameters can be interpreted as the returns to task content measures T_{jh} in the task pricing equations.

The effect of changes in π_{ht} on changes in the wage distribution are complex. To see this, consider the wage equation obtained by substituting equation (3) into (2):

$$w_{ijt} = \delta_t + c_j + Z_{it}\psi_t + \left[\pi_{0t} + \sum_{h=1}^{5} \pi_{ht} T_{jh} + \mu_{jt}\right] \sum_{k=1}^{K} \alpha_{jk} S_{ik}. \tag{4}$$

Since task prices and skills enter multiplicatively into the wage equation, a change in task prices linked to changes in the π_{ht} parameters has an impact on both the between- and

---

[4]Note that Yamaguchi assumes that the parameters α_{jk} are also functions of the task content variables T_{jh}, something we do not do since we would then need to be more specific about the way we introduce the K observed and unobserved skill components (corresponding of each parameter α_{jk}). More importantly, the question of whether or not the T_{jh}'s should be allowed to change over time in this setting is just a more structured way of thinking about the implications of possible changes in α_{jk}, an issue that we have already discussed.

within-group dimensions of inequality. For instance, even if the α_{jk} parameters were the same in all tasks/occupations, changes in π_{ht} would increase wage dispersion between occupations as long as average skills (e.g. education, one of the elements of the skill vector S_i) varied across occupations. Furthermore, since some dimensions of skills are unobserved, changes in π_{ht} also affect within-occupation inequality even after controlling for observable skills like education and experience.

Firpo, Fortin and Lemieux (2013) use this empirical model as a guide for carrying a full decomposition of overall changes in inequality. In this paper, we instead focus on the connection between the task content measures and changes in the between- and within-occupation wage dispersion. This is motivated by the fact that there are large differences in the changes in the level and dispersion of wages across occupations. This is illustrated in Figure 2 in the case of men over the 1990s. The figure shows the change in wages by decile (as a function of base period wages) in three broad occupation groups: food workers, skilled production workers, and engineers. In some "middle-end" occupations like production workers, all wage deciles decline in real terms, while they tend to increase in other occupations at the top-end (e.g. engineers) or low-end (e.g. food workers) of the distribution. Furthermore, wage dispersion increases for engineers (top wage deciles increase more than lower wage deciles) while the opposite happens for food workers (production workers are more neutral in this regard).

The main objective of the paper is to look at the connection between these wage changes and measures of the task content of occupations. With this in mind, we next introduce our key measures of task content based on the O*NET data.

3   Data

3.1   Occupational Measures of Technological Change and Off-shoring Potential

Like many recent papers (Goos and Manning (2007), Goos, Manning and Salomons (2009), Crino (2009)) that study the task content of jobs, and in particular their off-shorability potential, we use the O*NET data to compute our measures of technological change and offshoring potential.[5] Our aim is to produce indexes for all 3-digit occupations available in the CPS, a feat that neither Jensen and Kletzer (2007) nor Blinder

---

[5]Available from National Center for O*NET Development.

(2007) completed.[6] Our construction of an index of potential offshorability follows the pioneering work of Jensen and Kletzer (2007) [JK, thereafter] while incorporating some of the criticisms of Blinder (2007). The main concern of Blinder (2007) is the inability of the objective indexes to take into account two important criteria for non-offshorability: a) that a job needs to be performed at a specific U.S. location, and b) that the job requires face-to-face personal interactions with consumers. We thus pay particular attention to the "face-to-face" and "on-site" categories in the construction of our indexes.

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

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

We also consider five similar categories, but include five basic elements in each of these categories. Our first category "Information Content" regroups JK categories 1) and 2). It identifies occupations with high information content that are likely to be affected by ICT technologies; they are also likely to be offshored if there are no mitigating factor.[7] Appendix Figure 1 shows that average occupational wages in 2000-02 increase steadily with the information content. Our second category "Automation" is constructed using some work context measures to reflect the degree of potential automation of jobs

---

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

[7]Appendix Table 1 lists the exact reference number of the generalized work actitivies and work context items that make up the indexes.

and is similar in spirit to the manual routine index of Autor et al. (2003). The work context elements are: degree of automation, importance of repeating same tasks, structured versus unstructured work (reverse), pace determined by speed of equipment, and spend time making repetitive motions. The relationship between our automation index and average occupational wages display an inverse U-shaped left-of-center of the wage distribution. We think of these first two categories as being more closely linked to technological change, although 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, and that they capture the non-offshorability of jobs. 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. The relationship between these measures of offshorability (the reverse of non-offshorability) and average occupational wages are displayed in Appendix Figure 1. Automation and No-Face-to-Face contact exhibit a similar shape. No-Site is clearly U-shaped, and No-Decision-Making is steadily decreasing with average occupational wages.

For each occupation, O*NET provides information on the "importance" and "level" of required work activity and on the frequency of five categorical levels of work context.[8] 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.

---

[8]For example, the work context element "frequency of decision-making" has five categories: 1) never, 2) once a year or more but not every month, 3) once a month or more but not every week, 4) once a week or more but not every day, and 5) every day. The frequency corresponds to the percentage of workers in an occupation who answer a particular value. As shown in Appendix 1, 33 percent of sales manager answer 5) every day, while that percentage among computer programmers is 11 percent.

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

$$TC_{jh} = \sum_{k=1}^{A_h} I_{jk}^{2/3} L_{jk}^{1/3} + \sum_{l=1}^{C_h} F_{jl} * V_{jl}, \tag{5}$$

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

To summarize, we compute five different measures of task content using the O*NET: *i)* the information content of jobs, *ii)* the degree of automation of the job and whether it represents routine tasks, *iii)* the importance of face-to-face contact, *iv)* the need for on-site work, and *v)* the importance of decision making on the job. Call these five measures of task content (in each occupation j).

3.2 CPS Data

The empirical analysis is based on data for men from the 1973-78 May Supplements and 1979-2012 Outgoing Rotation Group (ORG) Supplements of the Current Population Survey. The data files were processed as in Lemieux (2006b) who provides detailed information on the relevant data issues. The wage measure used is an hourly wage measure computed by dividing earnings by hours of work for workers not paid by the hour. For workers paid by the hour, we use a direct measure of the hourly wage rate. CPS weights are used throughout the empirical analysis.

4 Empirical Test of the Occupational Wage Setting Model

4.1 Simple implications for means and standard deviations

We first discuss the implication of our wage setting model for the mean and standard deviation of occupational wages. Later in this section we expand the analysis to look at the whole distribution of occupational wages summarized using quantiles. To fix ideas, consider a simplified version of equation (2) where we ignore the covariates Z_{it}:

$$w_{ijt} = \delta_t + c_j + p_{jt}Y_{ij} \equiv \delta_t + c_j + p_{jt}\sum_{k=1}^{K} \alpha_{jk}S_{ik}. \tag{6}$$

As we discuss in Section 2, in this model changes in task prices p_{jt} have an impact on both the level and dispersion of wages across occupations. For instance, the average wage in occupation j at time t is

$$\overline{w}_{jt} = \delta_t + c_j + p_{jt}\overline{Y}_{jt}. \tag{7}$$

The standard deviations of wages is

$$\sigma_{jt} = p_{jt}\sigma_{Y,jt}, \tag{8}$$

where $\sigma_{Y,jt}$ is the standard deviation in tasks Y_{ij}, which in turns depends on the within-occupation distribution of skills S_{ik}. Since changes in both $\overline{w}_{jt}$ and σ_{jt} are positively related to changes in task prices p_{jt}, we expect these two changes to be correlated across occupations.

To see this more formally, assume that the within-occupation distribution of skills, S, and thus the distribution of task output, Y, remains constant over time (we discuss the assumption in more detail below). It follows that $\overline{Y}_{jt} = \overline{Y}_j$ and $\sigma_{Y,jt} = \sigma_{Y,j}$ for all t. Using a first order approximation of equations (7) and (8) and differencing yields:

$$\Delta\overline{w}_j \approx \Delta\delta + \overline{Y} \cdot \Delta p_j, \tag{9}$$

and

$$\Delta\sigma_j \approx \overline{\sigma_Y} \cdot \Delta p_j, \tag{10}$$

where $\overline{Y}$ ($\overline{\sigma_Y}$) is the average of $\overline{Y}_j$ ($\sigma_{Y,j}$) over all occupations j. Since the variation in Δp_j is the only source of variation in $\Delta\overline{w}_j$ and $\Delta\sigma_j$, the correlation between these two variables should be equal to one in this simplified model. In practice, we expect the correlation to be fairly large and positive, but not quite equal to one because of sampling error (in the estimates values of $\Delta\overline{w}_j$ and $\Delta\sigma_j$), approximation errors, etc.

A second implication of the model is that since task prices p_{jt} depend on the task content measures T_{jh} (see equation 3), these tasks content measures should help predict changes in task prices Δp_j, and thus $\Delta\overline{w}_j$ and $\Delta\sigma_j$. Differencing equation (3) over time we get:

$$\Delta p_j = \Delta\pi_0 + \sum_{h=1}^{5} \Delta\pi_h T_{jh} + \Delta\mu_j, \tag{11}$$

and, thus:

$$\Delta \overline{w}_j = \varphi_{w,0} + \sum_{h=1}^{5} \varphi_{w,h} T_{jh} + \xi_{w,h}, \qquad (12)$$

and

$$\Delta \sigma_j = \varphi_{\sigma,0} + \sum_{h=1}^{5} \varphi_{\sigma,h} T_{jh} + \xi_{\sigma,h}, \qquad (13)$$

where $\varphi_{w,0} = \Delta\delta + \overline{Y} \cdot \Delta\pi_0$; $\varphi_{w,h} = \overline{Y} \cdot \Delta\pi_h$; $\xi_{w,h} = \overline{Y} \cdot \Delta\mu_j$; $\varphi_{\sigma,0} = \overline{\sigma_Y} \cdot \Delta\pi_0$; $\varphi_{\sigma,h} = \overline{\sigma_Y} \cdot \Delta\pi_h$; $\xi_{\sigma,h} = \overline{\sigma_Y} \cdot \Delta\mu_j$. One important implication of the model highlighted here is that the coefficients $\varphi_{w,h}$ and $\varphi_{\sigma,h}$ should be proportional in equations (19) and (20) since they both depend on the same underlying coefficients $\Delta\pi_h$.

4.2 Empirical evidence for means and variances

We provide evidence that these two implications are supported in the data in the case of men in the 1990s. This group (and time period) is of particular interest since one important goal of this paper is to understand the sources of labor market polarization that was particularly important for that group/time period. Note that, despite our large samples based on three years of pooled CPS 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 here, we thus focus on occupations classified at the two-digit level (40 occupations) to have a large enough number of observations in each occupation.[9] All the estimates reported here (correlations and regression models) are weighted using the proportion of workers in the occupation.

The raw correlation between the changes in average wages and standard deviations is large and positive (0.44), as expected. It increases to 0.57 when we exclude agricultural occupations (less than three percent of the workforce).

We then run regression models for equations (19) and (20) using our five O*NET task content measures as explanatory variables. The regression results are reported in columns 1-4 of Table 1. Columns 1 and 2 show the estimated models for $\Delta \overline{w}_j$ and

---

[9]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).

$\Delta\sigma_j$, respectively, when all five task measure variables are included in the regression. The adjusted R-square of the regressions is equal to 0.52 for both models, indicating that our task content measures capture most of the variation in changes in the level $(\Delta\overline{w}_j)$ and dispersion $(\Delta\sigma_j)$ of wages over occupations. Since several of the coefficients are imprecisely estimated, we also report in columns 3 and 4 estimates from separate regressions for each task content measure. The task content measures are significant in most cases, and the sign of the coefficient estimates are the same in the models for changes in average wages and standard deviations. This strongly support the prediction of our wage setting models that the estimated effect of the task content measures should be proportional in the models for average wages and standard deviations.

Note also that, in most of the cases, the sign of the coefficients conforms to expectations. 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 changes in task prices. 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 a negative effect for the "automation/routine" measure. 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 with a high level of "decision-making", to do relatively well in the presence of offshoring. Since these last three variable capture non-offshorability, they are entered as their reverse in the regression and we should expect their effect to be negative.

In columns 3 and 4, all the estimated coefficients are of the expected sign except for the "no onsite" task. This may indicate 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 whose job could be offshored), vs. working on a site in the United States (i.e. a construction worker).

One potential issue with these estimates is that we are only using the raw changes in $\overline{w}_{jt}$ and σ_{jt} that are unadjusted for differences in education and other characteristics. Part of the changes in $\overline{w}_{jt}$ and σ_{jt} may, thus, be due to composition effects or changes in the return to underlying characteristics (like education) that are differently distributed across occupations. To control for these confounding factors, we reweight the data using simple logits to assign the same distribution of characteristics to each of the 40 occupations in the two time periods.[10]

---

[10]We use a set of five education dummies, nine experience dummies, and dummies for marital status

This procedure allows us to relax the assumption that the distribution of skills S is constant over time. Strictly speaking, we can only adjust for observable skills like education and experience. To deal with unobservables, we could then invoke an ignorability assumption to ensure that, conditional on observable skills, the distribution of unobservable skills is constant over time. A more conservative approach is to view the specifications where we control for observable skills as a robustness check.

The results reported in columns 5-8 indeed suggest that the main findings discussed above are robust to controlling for observables. Generally speaking, the estimated coefficients have similar magnitudes and almost never change sign relative to the models reported in column 1-4. Overall, the results presented here strongly support the predictions of our wage setting model.

4.3 Quantiles of the occupational wage structure

One disadvantage of using the standard deviation (or the variance) as a measure of wage dispersion is that it fails to capture the polarization of the wage distribution that has occured since the late 1980s. As a result, we need an alternative way of summarizing changes in the wage distribution for each occupation that is yet flexible enough to allow for different changes in different parts of the distribution. We do so by estimating linear regression models for the changes in wages at different quantiles of the wage distribution for each occupation. As we now explain in more detail, the intercept and the slope from these regressions are the two summary statistics we use to characterizes the changes in the wage distribution for each occupation.

We now extend or appraoch by looking at all quantiles of the of the wage distribution for each occupation. Consider $F_{jt}(w)$, the distribution of effective skills $\sum_{k=1}^{K} \alpha_{jk} S_{ik}$ provided by workers in occupation j at time t. Under the admittedly strong assumption that the distribution of skills supplied to each occupation is stable over time, we can write the qth quantile of the distribution of wages in occupation j at time t as:

$$w_{jt}^q = \overline{w}_{jt} + p_{jt} F_j^{-1}(q). \tag{14}$$

---

and race as explanatory variables in the logits. The estimates are used to construct reweighting factors that are used to make the distribution of characteristics in each occupation-year the same as in the overall sample for all occupations (and time periods).

Taking differences over time yields

$$\Delta w_j^q = \Delta \overline{w}_j + \Delta p_j F_j^{-1}(q). \tag{15}$$

Solving for $F_j^{-1}(q)$ in equation (14) at the base period $t = 0$, and substituting into equation (15) yields

$$\Delta w_j^q = \Delta \overline{w}_j - \frac{\Delta p_j}{p_{j0}} \overline{w}_{j0} + \frac{\Delta p_j}{p_{j0}} w_{j0}^q, \tag{16}$$

or

$$\Delta w_j^q = a_j + b_j w_{j0}^q, \tag{17}$$

where $a_j = \Delta \overline{w}_j - \frac{\Delta p_j}{p_{j0}} \overline{w}_{j0}$ and $b_j = \frac{\Delta p_j}{p_{j0}}$.

Interestingly, the coefficient on the base period wage quantile w_{j0}^q is simply the change in the task price p_{jt} expressed in relative terms. This suggests a very simple way of estimating relative changes in task prices in each occupation. First compute a set of wage quantiles for each occupation in a base and an end period. Then simply run a regression of changes in quantiles on base period quantiles. The slope coefficient of the regression, b_j, provides a direct estimate of the relative change in task price, $\frac{\Delta p_j}{p_{j0}}$.

Our simple wage setting model is highly parametrized since changes in wages in a given occupation is only allowed to depend on task prices p_{jt}. While this parsimonious specification provides a simple interpretation for changes in occupational wages, actual wage changes likely depend on other factors. 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 can be captured by allowing for a percentile-specific component λ^q which leads to the main regression equation to be estimated in the first step of the empirical analysis:

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

where we have also added an error error term ε_j^q to capture other possible, but unsystematic, departures from our simple task pricing model.

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

14

4.3.1 Connecting to occupational task measures

In the second step of the analysis, we link the estimated intercepts and $(a_j$ and $b_j)$ to measures of the task content of each occupation, as we did in the case of the mean and standard deviation earlier.

The second step regressions are

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

and

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

4.4 Occupation Wage Profiles: Results

We now present the estimates of the linear regression models for within-occupation quantiles (equation (18)), and then link the estimated slope and intercept parameters to our measures of task content from the O*Net. We refer to these regressions as "occupation wage profiles".

As in the analysis for means and standard deviations, the empirical analysis is based on data for men from the 1988-90 and 2000-02 Outgoing Rotation Group (ORG) Supplements of the Current Population Survey. The main reason for focusing this first part of the analysis on the 1990s is that it represents the time period when most of the polarization of wages phenomenon documented by Autor, Katz and Kearney (2006) occurred. The choice of years is also driven by data consistency issues since there is a major change in occupation coding in 2003 when the CPS switches to the 2000 Census occupation classification. This makes it harder to compare detailed occupations from the 1980s or 1990s to those in the post-2002 data.

Before presenting our main estimates, consider again the overall changes in the wage distribution illustrated in Figure 1. Consistent with Autor, Katz and Kearney (2006), Figure 1a shows that 1988-90 to 2000-02 changes in real wages at each percentile of the male 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.[11] The figure shows that wage changes in the top half of the

---

[11]Given the major change in occupational classification after 2002, we did not attempt to reconcile the coding of occupations before and after this change. This explains why we only compare post-2002 years (2003-2004 to 2009-2010) when documenting more recent changes in the wage distribution.

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.[12] As a reference, we also present in Figure1b the same descriptive statistics for women that are qualitatively similar to those for men.

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.[13] 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.[14]

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

[12]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.

[13]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).

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

observations (weighted using the earnings weight from the CPS) in each occupation.

Figure 3a presents the raw data used in the analysis. The figure plots the 360 observed changes in wages (9 observation for each of the 40 occupations) as a function of the base wages. The most noticeable feature of Figure 3a is that wage changes exhibit the well-known U-shaped pattern documented by Autor, Katz, and Kearney (2006) that we also see in Figure 1a. Broadly speaking, the goal of the first part of the empirical analysis is to see whether the simple linear model presented in equation (18) helps explain a substantial part of the variation documented in Figure 3a.

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

The second measure of fit consists of looking at whether the model is able to explain the U-shaped feature of the raw data presented in Figure 3a. As a reference, the estimated coefficient on the quadratic term in the fitted (quadratic) regression reported in Figure 3a is equal to 0.136. For each estimated model, we run a simple regression of the regression residuals on a linear and quadratic term in the base wage to see whether there is any curvature left in the residuals that the model is unable to explain.

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

As in the case of the means and variances, we address these issues by reweighting the distribution of covariates in each occupation at each time period so that it is the same as in the pooled distribution with all occupations and time periods (1988-90 and 2000-02). This involves computing 80 separate logits (40 occupations times two years) to perform a DiNardo, Fortin, and Lemieux (1996) reweighting exercise. The various quantiles of

the wage distribution for each occupation are then computed in the reweighted samples. The covariates used in the logits are a set of five education dummies, nine experience dummies, and dummies for race and marital status. The unadjusted models are reported in Panel A of Table 2, while the estimates that adjust for the covariates by reweighting are reported in Panel B. Since the results with and without the adjustment are qualitatively similar, we focus our discussion on the unadjusted estimates reported in Panel A.

As a benchmark, we report in column 1 the estimates from a simple model where the only explanatory variable is the base wage. This model explains essentially none of the variation in the data as the adjusted R-square is only equal to 0.0218. This reflects the fact that running a linear regression on the data reported in Figure 3a essentially yields a flat line. Since the linear regression cannot, by definition, explain any of the curvature of the changes in wages, the curvature parameter in the residuals (0.136) is exactly the same as in the simple quadratic regression discussed above.

In column 2, we only include the set of occupation dummies (the a_j's) in the regression. The restriction imbedded in this model is that all the wage deciles within a given occupation increase at the same rate, i.e. there is no change in within-occupation wage dispersion. Just including the occupation dummies explains 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.

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

By contrast, the fit of the model improves drastically once we introduce occupation-specific slopes in column 5. The R-square of the model jumps to 0.9274, 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 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

18

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

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

We next illustrate the fit of the model by plotting occupation-specific regressions for the 30 largest occupations curves in Figure 3b.[15] 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. This is consistent with the model of Section 2 where the 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 explore this hypothesis more formally by estimating the regression models in equations (19) and (20) that link the intercept and slopes of the occupation wage change profiles to the task content of occupations.[16] The results are reported in Table 3. In the first four columns of Table 3, 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.

To get a better sense of how these task measures vary across the occupation distribution, consider again Appendix Figure A1, which plots the values of the task index as a function of the average wage in the (3-digit) occupation. The "information content" and

---

[15]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).

[16]To be consistent with equation (**??**), 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.

"decision making" measures are strongly positively related to wages. Consistent with Autor, Levy and Murnane (2003), the "automation" task follows an inverse U-shaped curve. To the extent that technological change allows firms to replace workers performing these types of tasks with computer driven technologies, we would expect both the intercept and slope of occupations with high degree of automation to decline over time.

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

The strongest and most robust result in Table 3 is that occupations with high level of automation experience a relative decline in both the intercept and the slope of their occupational wage profiles. The effect is statistically significant in six of the eight specifications reported in Table 3. 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. For instance, the returns to skill in jobs that do not require face-to-face contacts will likely decrease since it is now possible to offshore these types of jobs to another country.

The results reported in Table 3 are mixed. As expected, the effect of "no face to face" and "no decision making" is generally negative. By contrast, the effect of "no on-site work" is generally positive, which is surprising. One possible explanation is that the O*NET is not well suited for distinguishing whether a worker has to work on "any site" (i.e. an assembly line worker), vs. working on a site in the United States (i.e. a construction worker).

On balance, most of the results reported in Table 3 are consistent with our expecta-

tions. More importantly, 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.

We draw two main conclusions from Table 3. First, as predicted by the linear skill pricing model of Section 2, the measures of task content of jobs tend to have a similar impact on the intercept and the slope of the occupational profiles. Second, tasks account for a large 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.

5 Conclusion

In this paper, we look at the contribution of occupations to changes in the distribution of wages. We present a simple model of skills, tasks, and wages, and use this as a motivation for estimating models for the change means, variances, and occupation-specific wage percentiles between 1988-90 and 2000-02. The findings suggest that changes in occupational wage profiles help explain the U-shaped of changes in the wage distribution over this period. We also find that measures of technological change and offshoring at the occupation level help predict the changes in the occupational wage profiles.

REFERENCES

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

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

AUTOR, D.H., KATZ, L.F. AND M.S. KEARNEY (2006), "The Polarization of the U.S. Labor Market," *American Economic Review* 96, 189–194.

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

BARSKY, R., J. BOUND, K. CHARLES, AND J. LUPTON (2002), "Accounting for the Black-White Wealth Gap: A Nonparametric Approach," *Journal of the American Statistical Association*, 97(459), 663-673.

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

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

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

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

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

FIRPO, S. FORTIN, N.M., AND T. LEMIEUX (2007) "Decomposing Distribution Using Recentered Influence Fucntion Regressions," Unpublished manuscript, PUC-Rio and UBC

FIRPO, S. FORTIN, N.M., AND T. LEMIEUX (2009) "Unconditional Quantile Regressions," *Econometrica* 77(3), May, 953-973.

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

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

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

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

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

GIBBONS, R., L.F. KATZ, T. LEMIEUX, AND D. PARENT "Comparative Advantage, Learning, and Sectoral Wage Determination," *Journal of Labor Economics* 23(4), October, 681-724

GOOS, MAARTEN AND ALAN MANNING (2007) "Lousy and Lovely Jobs: The Rising Polarization of Work in Britain", *Review of Economics and Statistics*, 89(1), 118-133

GOOS, MAARTEN, ALAN MANNING, AND ANNA SALOMONS (2009) "Recent Changes in the European Employment Structure: The Roles of Technological Change, Globalization and Institutions," unpublished manauscript.

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

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

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

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

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

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

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

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

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

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

Figure 1. Changes in Real Log Wages by Percentile


A. Men

B. Women
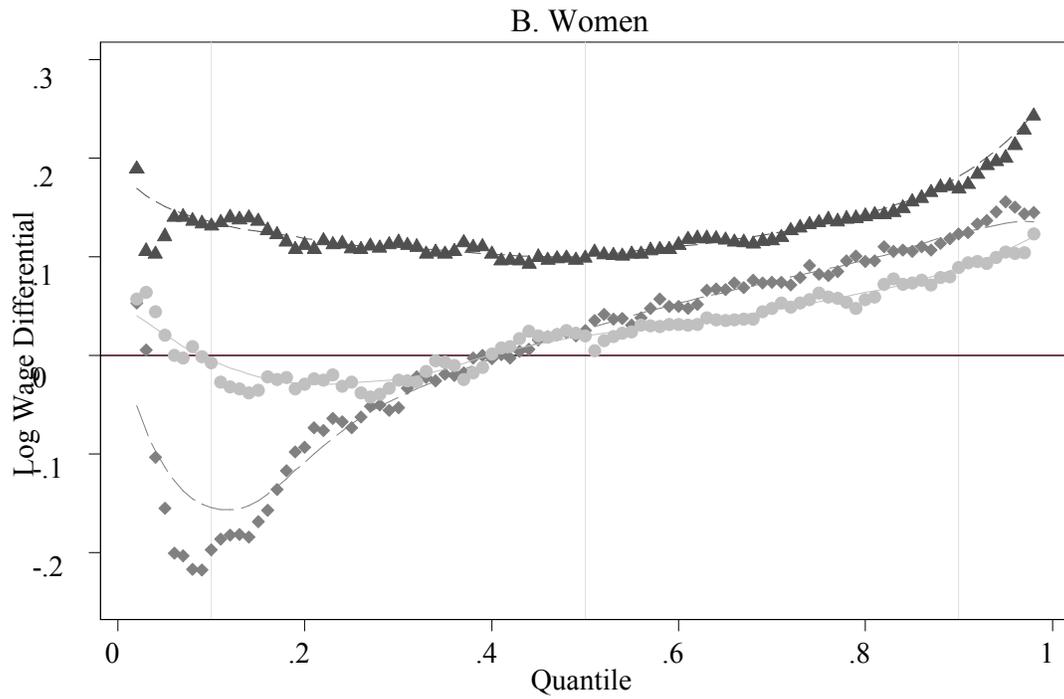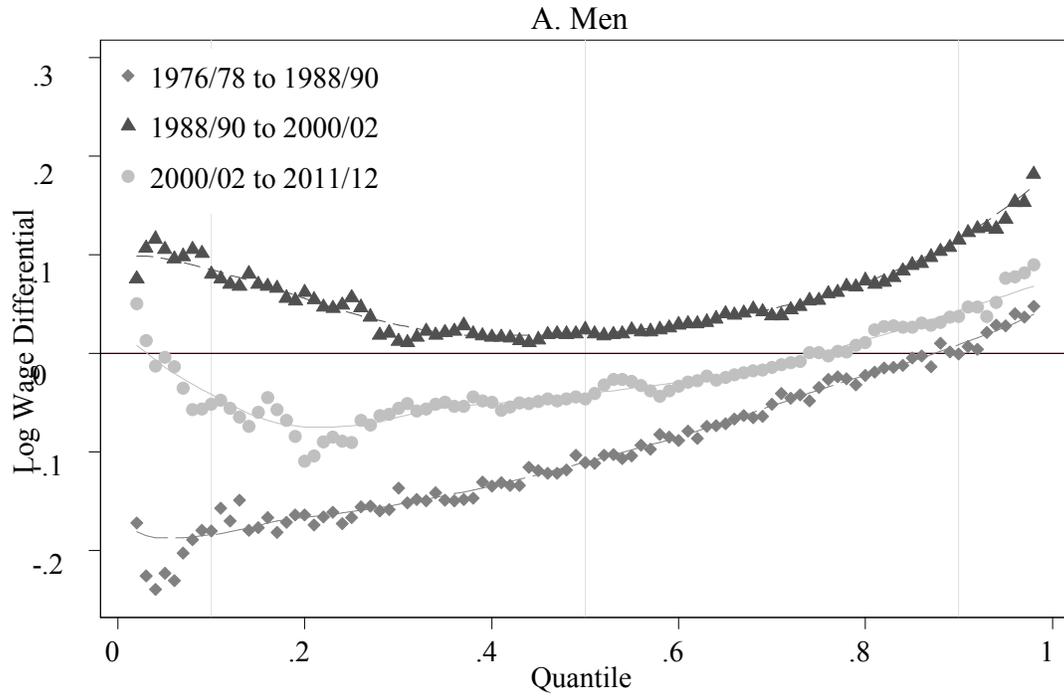
Figure 2. Changes in Within and Between Occupations Wage by Decile

Selected Occupations - Men 1988/90 to 2000/02

Food Services

Engineers

Skilled Production Workers

Wage change: Δw

Base period wage: w_0^q
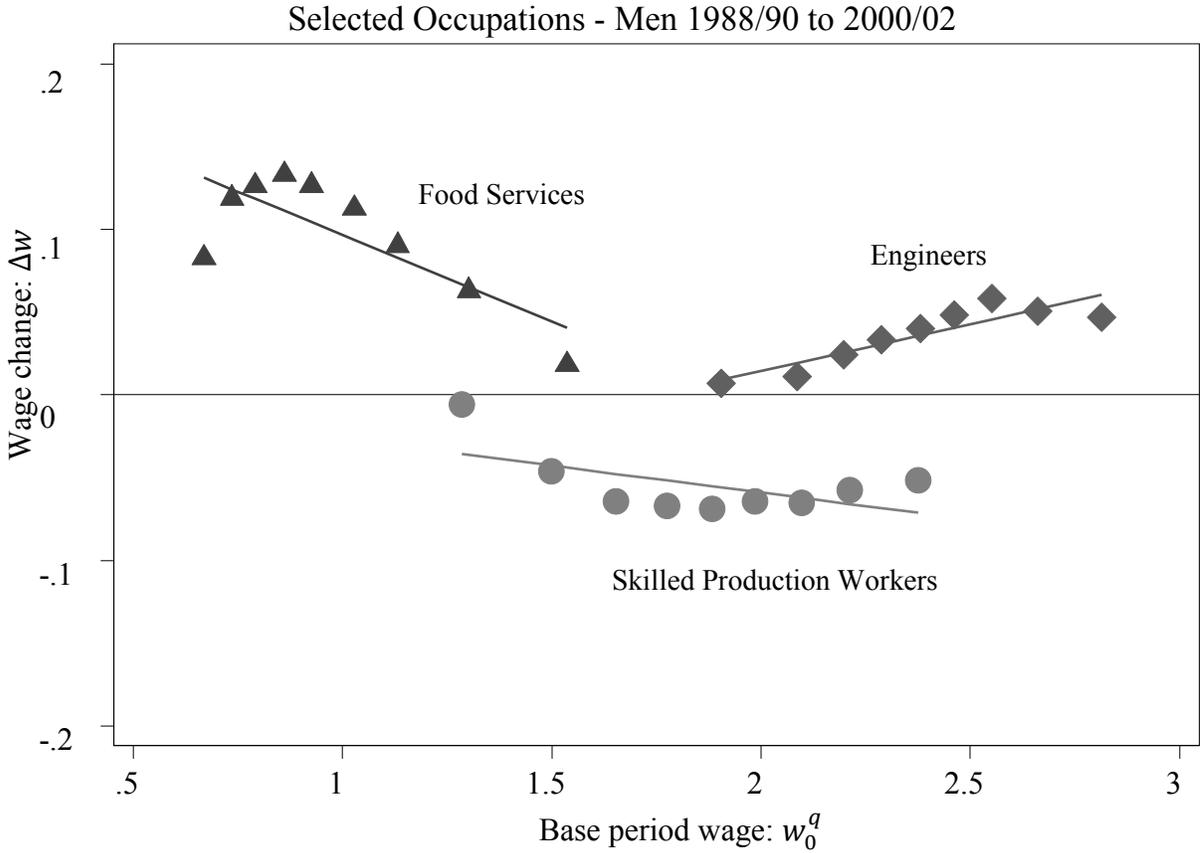
Figure 3a: Change in wage by 2-digit occupation
1983-85 to 2000-02 change for each decile
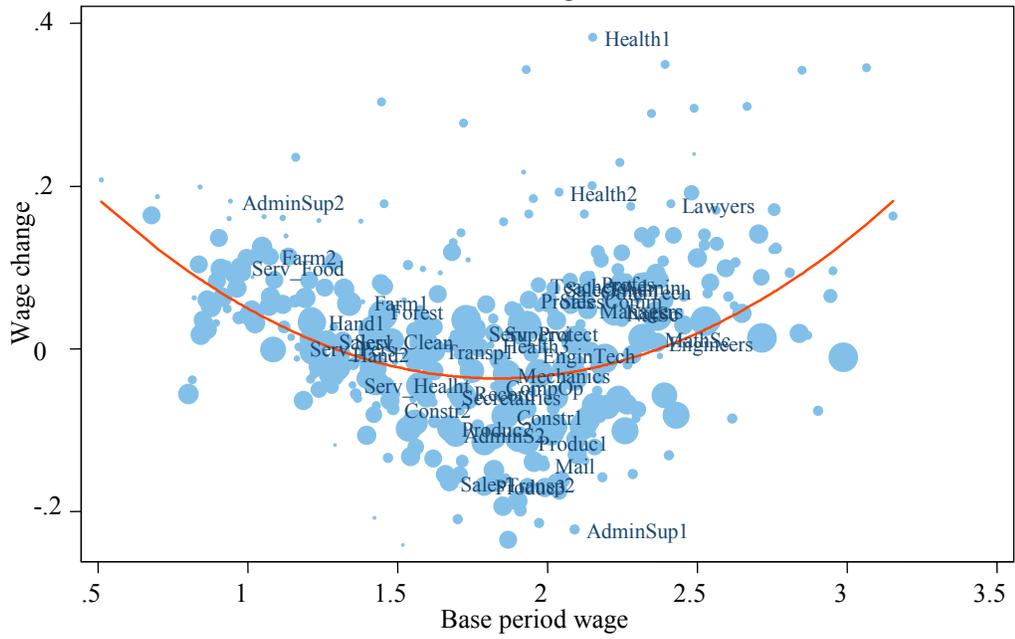
Figure 3b: Fitted change in wage by 2-digit occupation
### 1983-85 to 2000-02 change for each decile
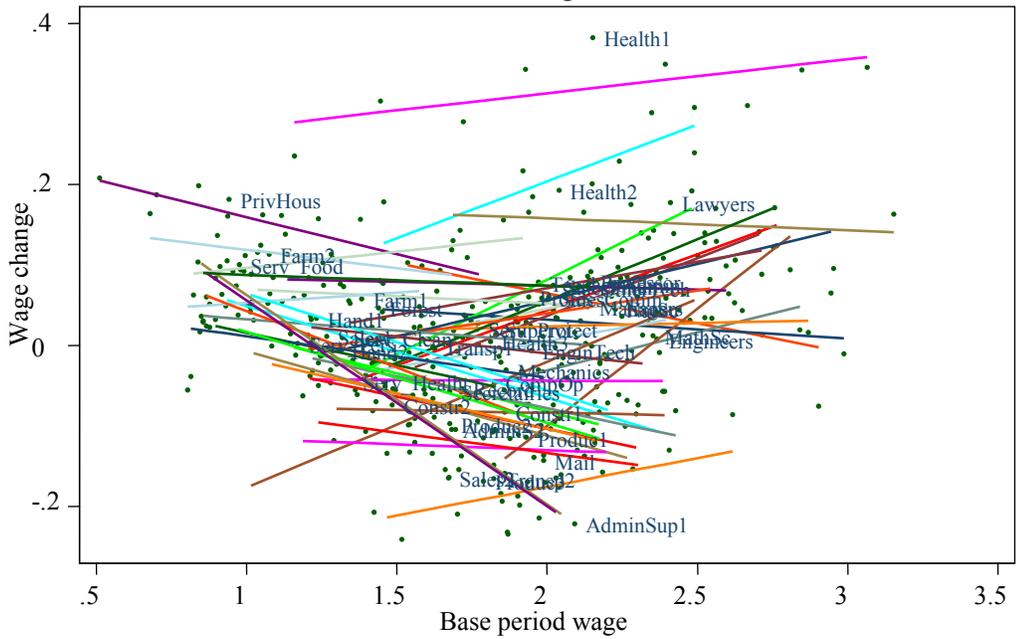
Table 1. Estimated Effect of Task Requirements on Average Wages and Standard Deviations
Men, 1988-90 to 2000-02, 2-digit Occupations

Tasks entered:	Together		Separately		Together		Separately	
Dep. variable:	Average	Std dev	Average	Std dev	Average	Std dev	Average	Std dev
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Raw changes				Reweighted changes			
Information	0.0106	0.0041	0.0198***	0.0163***	0.0081***	0.0058	0.0081	0.0179***
content	(0.0108)	(0.0059)	(0.0079)	(0.0039)	(0.0096)	(0.0055)	(0.0061)	(0.0037)
Automation	-0.0306***	-0.0096	-0.0467***	-0.0226***	-0.0137	-0.0146**	-0.0228***	-0.0245***
/routine	(0.0112)	(0.0062)	(0.0089)	(0.0053)	(0.01)	(0.0057)	(0.0078)	(0.0051)
No on-site work	0.0018	0.0099***	0.0190***	0.0132***	0.0021	0.0077**	0.0118***	0.0118***
	(0.0059)	(0.0033)	(0.0048)	(0.0023)	(0.0052)	(0.003)	(0.0037)	(0.0025)
No face-to-face	-0.0418***	0.0082	-0.0560***	-0.0207	-0.0348**	0.0165**	-0.0320**	-0.0180**
	(0.0148)	(0.0081)	(0.0109)	(0.0069)	(0.0131)	(0.0075)	(0.0088)	(0.007)
No decision	0.0229***	-0.0077	-0.0275***	-0.0189***	0.00227*	-0.0125**	-0.0079	0.0221***
making	(0.0163)	(0.009)	(0.0105)	(0.0056)	(0.0144)	(0.0083)	(0.0084)	(0.0052)
Adj. R-square	0.517	0.523	---	---	0.306	0.582	---	---

Notes: All models are estimated by running regressions of the occupation-specific changes in average wages and standard deviations on the task content measures. The models reported in all columns are weighted using the fraction of observations in each occupation in the base period (1988-90). In columns 5-8 the data are reweighted so that the distribution of characteristics in each occupation and time period is the same as in the overall sample (for both periods pooled). See the text for more detail.

Table 2: 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)
A. Models without controls for observables						
Adj. R-square	0.0218	0.5535	0.0284	0.5996	0.9274	0.8602
Curvature	0.135	0.087	0.073	0.073	0.002	0.020
in residuals	(0.009)	(0.005)	(0.006)	(0.006)	(0.003)	(0.004)
B. Models with controls for observables						
Adj. R-square	0.0730	0.2525	0.0498	0.4644	0.8743	0.7711
Curvature	0.131	0.116	0.084	0.068	0.002	0.052
in residuals	(0.015)	(0.011)	(0.015)	(0.011)	(0.005)	(0.007)
Occupation dummies		X		X	X	X
Decile dummies			X	X	X	
Base wage	X				X	X
Occ * base wage					X	X

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

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

	Tasks included together						Included separately	
	Intercept			Slope			Intercept	Slope
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Information content	0.005	0.030	-0.001	0.020	-0.008	0.010	0.035	0.031
	(0.011)	(0.011)	(0.012)	(0.012)	(0.013)	(0.024)	(0.009)	(0.013)
Automation /routine	-0.016	-0.030	-0.018	-0.028	-0.013	-0.017	-0.034	-0.045
	(0.011)	(0.010)	(0.011)	(0.013)	(0.012)	(0.022)	(0.008)	(0.010)
No on-site work	0.003	0.002	0.003	0.019	0.021	0.020	0.015	0.026
	(0.006)	(0.005)	(0.006)	(0.007)	(0.006)	(0.011)	(0.004)	(0.004)
No face-to-face	-0.036	0.002	0.000	0.027	-0.014	-0.021	-0.035	-0.051
	(0.015)	(0.015)	(0.017)	(0.017)	(0.018)	(0.034)	(0.009)	(0.011)
No decision making	0.032	0.001	-0.006	-0.045	-0.012	0.026	-0.035	-0.051
	(0.017)	(0.015)	(0.017)	(0.019)	(0.018)	(0.035)	(0.012)	(0.015)
Base wage	no	yes	yes	no	yes	yes	yes	yes
Reweighted	no	no	yes	no	no	yes	no	no
Adj. R-square	0.27	0.51	0.38	0.73	0.81	0.51	---	---

Notes: All models are estimated by running regressions of the 40 occupation-specific intercepts and slopes (estimated in

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