The Inequality Implications of Occupational Evolution

Enghin Atalay, Phai Phongthiengtham, Sebastian Sotelo, Daniel Tannenbaum*

June 2019

Please click here for most recent version

Abstract

This paper explores the inequality implications of the evolution of occupations. It develops an equilibrium model of occupational choice, based on worker's comparative advantage, in which jobs are bundles of tasks that a worker must perform. We bring this model to measures on the changing task contents of within occupations, and we find that shifts in the relative demand for tasks account for a large fraction of the increase in 90-10 earnings inequality observed over our sample period. **JEL Codes: E24, J20, O33**

1 Introduction

Labor income inequality in the United States has increased considerably over the last several decades. Between 1960 and 2000, the ratio of earnings for the median worker, compared to the worker at the 10th percentile of the earnings distribution, has increased from 2.75 to 2.86. Over the same period, the 90-50 earnings ratio increased even more starkly, from 1.81 to 2.27. Alongside rising inequality, there have been dramatic changes in the composition of

*Atalay and Phongthiengtham: Department of Economics, University of Wisconsin-Madison. Sotelo: Department of Economics, University of Michigan-Ann Arbor. Tannenbaum: Department of Economics, University of Nebraska-Lincoln. We thank Sara Heller, Andrei Levchenko, Pablo Ottonello, Pascual Restrepo, and Chris Taber for constructive and helpful comments, and Erin Robertson for excellent editorial assistance. We also thank seminar audiences at APPAM, ASSA, Cornell, Geneva, Iowa State, Lima, Michigan, Penn State, the NBER Summer Institute, the University of British Columbia, the Upjohn Institute, Virginia Tech, Washington University, and Wisconsin. We acknowledge financial support from the Washington Center for Equitable Growth. This work is also being supported by Grant #92-18-05 from the Russell Sage Foundation.

U.S. employment. For example, the employment share of occupations intensive in routine tasks has shrunk, while the share of occupations emphasizing nonroutine tasks, as well as social and cognitive skills, has grown (Autor, Levy, and Murnane, 2003; Autor, Katz, and Kearney, 2005; Autor and Dorn, 2013; and Deming, 2017).

In this paper, we quantitatively explore the inequality implications of these within-occupation shifts in task composition. To do so, we build an equilibrium model of worker sorting across occupations and map it to the new data in Atalay, Phongthiengtham, Sotelo, and Tannenbaum (Forthcoming). In our model, individual occupations are represented as a bundle of tasks. Workers' skills govern their abilities to perform each of the individual tasks in their occupation, and give rise to comparative advantage. These skill levels are functions of workers' observable characteristics — like gender, education, and experience — but also contain an idiosyncratic component. Based on their skill levels and the demand for tasks within each occupation, workers select into the occupation with the highest payoff.

Using the text of newspaper vacancy ads, Atalay, Phongthiengtham, Sotelo, and Tannenbaum (Forthcoming) introduce a new dataset to study the evolution of jobs in the U.S. between 1940 and 2000. The data show a pervasive move towards nonroutine tasks across occupations. It also suggests that, in accounting for the aggregate decline of routine tasks and the rise of nonroutine tasks, the contribution of changes in task composition within occupations is at least comparable to that employment shifts between jobs.

We estimate each demographic group's skills to perform different tasks, combining information on demographic groups' earnings and occupation choices. Using our estimated model, we calculate that changes in the relative demands for tasks have led to a 16 log point increase in 90-10 inequality among men, and a 23 log point increase in inequality for men and women combined. The intuition is that workers who are at the bottom of the earnings distribution have a comparative advantage in occupations that are intensive in manual tasks, while the demand for these tasks has declined.

Our paper contributes to several strands of literature. The first shows how the aggregate demand for tasks has evolved and the implications of this evolution for inequality. A main building block of this literature is the hypothesis that, due to technological advances and trade, the aggregate demand for routine tasks has declined, which in turn reduced the wages of workers who have a comparative advantage in those tasks (see, for example, Autor, Levy, and Murnane, 2003, Acemoglu and Autor, 2011, and Michaels, Rauch, and Redding (2016)). This literature has also shown that, while routine-intensive jobs have declined, employment and wages jobs intensive in social and cognitive skills has been on the rise (Deming, 2017). Burstein, Morales, and Vogel (2015) and Michaels, Rauch, and Redding (2016) use quantitative frameworks to measure the impact of new technologies on between-group wage inequality and the specialization of cities in interaction-intensive sectors. We contribute to this literature by exploring quantitatively the impact of task changes within occupations between 1960 and 2000. The impact of this margin of change has not been measured before, and, in fact, was not observable on a continual basis before Atalay, Phongthiengtham, Sotelo, and Tannenbaum (Forthcoming) introduced the new data set we use in this paper. We show that changes in the task contents of occupations can account for a substantial share of the inequality increase in the U.S. during the past decades.

A previous approach to measure the impact of task demand on inequality is the one based on RIF regressions, advanced among others, by Firpo, Fortin, and Lemieux (2014). This approach decomposes changes in the distribution of wages into the contribution of occupational characteristics and other factors (including de-unionization, changes in minimum wage, and changes in worker demographics). We employ an approach that we belive is highly complementary to one based on RIF regressions. While we make particular distributional assumptions, our framework is well suited to handle sorting and reallocation in response to particular counterfactual shocks.

To study the evolution of individual occupations, we build on Lagakos and Waugh (2013), Burstein, Morales, and Vogel (2015), and Galle, Rodriguez-Clare, and Yi (2017). Relative to this literature, we innovate in modeling occupations as a bundle of tasks that workers need to perform. We allow for the importance of these tasks to change over time, reflecting what we observe in the data. Because we estimate the skill of workers of different groups to perform these tasks, our model has an observable, time-varying margin of comparative advantage not present in previous models. The key takeaway from our analysis is that keeping everything else constant, the observed evolution of occupations, interpreted as a demand shock, generates an increase in inequality of the same magnitude as the overall observed increase between 1960 and 2000.

The rest of the paper is organized as follows. Section 2 describes the data we use in our quantification. Section 3 introduces a model of worker sorting across occupations and Section 4 discusses how we quantify this model and use it to account for inequality in equilibrium. Section 5 reviews our results and suggests areas for future research.

2 Data

In this Section we discuss the two main data sources we use in the paper. First, we briefly summarize the dataset in Atalay, Phongthiengtham, Sotelo, and Tannenbaum (Forthcoming). Second, we discuss the measures of employment and wages we obtain from the Census.

Occupational Evolution Data. To measure changes within occupations over time, we use the data introduced by Atalay, Phongthiengtham, Sotelo, and Tannenbaum (Forthcoming). The dataset computes, for each year and occupation, the frequency of words that map to a standard task classification coming from Spitz-Oener (2006), which divides tasks into

Routine (Manual and Cognitive) and Nonroutine (Manual, Interactive, and Analytic). We denote this frequency by $\tilde{T}_{hjt}$ in the rest of the paper.[1]

The dataset provides these measures at different levels of aggregation. Since we combine these data with wages coming from the Census (see below) we rely on their mapping between job titles and occupation codes, which yield measurements at the 4-digit SOC codes.

Employment and Earnings Data. We also use micro data from U.S. Census for 1960, 1970, 1980, 1990, and 2000 (Ruggles, Genadek, Goeken, Grover, and Sobek, 2015). We aggregate individual data to cells defined by gender, education, experience, and occupational code. For each cell, we calculate employment and average earnings.

We draw from the sample of full time workers—ages 16 to 65, working for wages, having worked at least 40 weeks in the previous year—whose gender, age, occupation, and education data are not imputed.[2] With this sample, employment shares reflect numbers of workers in each SOC occupation, in each decennial census year.

3 A Model of Worker Sorting and Task Contents

In this section, we explore the implications of changes in occupations' task content for the earnings distribution. To do so, we develop a framework for interpreting occupations as a bundle of tasks. We embed this framework into a quantitative general equilibrium model of occupational sorting based on comparative advantage (akin to Heckman and Sedlacek, 1985, Heckman and Scheinkman, 1987, or more recently Burstein, Morales, and Vogel, 2015). In the next Section, we estimate this model and then use it to quantify the impact of changes in the demand for tasks on earnings inequality.

3.1 Occupations as bundles of tasks

In this section we take an alternative approach to calculating changes in the earnings distribution. We begin with a quantitative model in which workers sort based on comparative advantage. We consider the effect of changes in the parameters of occupations' production functions, and interpret these changes as reflecting exogenous shifts in the demand for worker-produced tasks.[3]

[1] These data are available at https://occupationdata.github.io/. The site contains data at different levels of aggregation and for additional job characteristics.

[2] To map SOC codes to Census occ1990 codes, we construct our own crosswalk. The details of this crosswalk are contained in (Atalay, Phongthiengtham, Sotelo, and Tannenbaum, Forthcoming), footnote 20.

[3] We have subsequently extended this framework to evaluate the impact of the arrival of Information and Communication Technologies on inequality in Atalay, Phongthiengtham, Sotelo, and Tannenbaum (2018).

Environment and Endowments We consider a sequence of static closed economies in periods $t = 1960, 1970, ..., 2000$. In each period there are a continuum of workers, indexed by i. We group workers into a finite number of groups, indexed by $g = 1, \ldots, G$. The mass of workers in group g is L_{gt}.

Technology There is a set of occupations, $j = 1, \ldots, J$ into which workers sort according to their comparative advantage. Workers of type g are endowed with a set of skills S_{gh}, $h = 1, ..., H$, where h indexes tasks that workers perform in their occupations. Worker i produces task h by combining skills and time: $q_{iht} = S_{gh} \cdot l_{iht}$. Here, l_{iht} equals the amount of time the worker allocates to task h. Each worker has a total of one unit of time, which she supplies inelastically in one of $j = 1, ..., J$ occupations.

Worker i, if she works in occupation j, produces

$$V_{ijt} = \epsilon_{ijt} \prod_{h=1}^{H} \left(\frac{q_{iht}}{T_{hjt}} \right)^{T_{hjt}} \tag{1}$$

units of occupation-specific output. In Equation 1, ϵ_{ijt} is an occupation-worker unobserved efficiency level, and T_{hjt} gives the importance of task h in occupation j. We fix $\sum_h T_{hjt} = 1$ for each occupation and time period.[4] Variation in T_{hjt} across occupations captures, for example, that some occupations are more intensive in routine analytic tasks than others. To match our time-varying measures of task content, we allow the parameters T_{hjt} to evolve exogenously through time. Such changes could reflect, again as an example, declines in the demand for worker-performed routine manual or routine cognitive tasks caused by a decline in the price of computers, or increases in nonroutine interactive tasks related to the increased importance of team management.

Preferences To close the model, we assume that a representative household has CES preferences over the output produced by the different occupations:

$$V_t = \left[\sum_{j=1}^{J} (\xi_j)^{\frac{1}{\rho}} \cdot (V_{jt})^{\frac{\rho-1}{\rho}} \right]^{\frac{\rho}{\rho-1}},$$

with V_{jt} representing the sum of the output of individual workers, i, who produce occupation j output. Here, ξ_j parameterizes the consumer's preferences for occupation j output, and ρ gives the preference elasticity of substitution across occupations' output.

[4]Throughout the model and its estimation, we adopt the stance that the five Spitz-Oener task categories proxy for the full set of mutually exclusive occupational tasks. A potentially omitted sixth task category would affect the estimation and results insofar as they represent a large share of worker time, and are orthogonal to — in their trends and skill requirements — those in the observed task groups.

3.2 Equilibrium

We assume that worker i's wages, W_{ijt}, equal the value of the output she produces, $P_{jt}V_{ijt}$, which represents the product of the output produced and the price of the occupation j specific output. Taking P_{jt} as given, each worker chooses the amount of time spent on each of the H tasks to maximize the value of her output. The solution to this problem implies that $l_{ijt} = T_{hjt} \cdot \left(\sum_{h'=1}^{H} T_{h'jt}\right)^{-1} = T_{hjt}$. Due to the assumption that the production function of occupation-j specific output is Cobb-Douglas, individual abilities do not shape the time spent in each task, although they will affect occupational choice.

Plugging the optimal time allocation back into Equation 1 yields worker i's log wages

$$\log W_{ijt} = \log P_{jt} + \sum_{h=1}^{H} T_{hjt} \log S_{gh} + \log \epsilon_{ijt}. \tag{2}$$

We assume that workers are freely able to choose their occupations to maximize their wages. Workers take as given each occupation-specific output price, and the task activities associated with each occupation. The idiosyncratic component of worker i's returns from working in occupation j, ϵ, is a Frechet-distributed random variable, drawn independently across occupations and workers, with shape parameter θ, i.e., $\Pr\left[\epsilon_{ijt} < z\right] = \exp\left(-z^{-\theta}\right)$. The parameter θ is an inverse measure of the dispersion of ϵ_{ijt} within group-occupation pairs.

The Frechet formulation conveniently generates expressions for the fraction of group g workers who work in occupation j and the average wage of group g workers. The probability that a worker of type g sorts into occupation j is

$$\lambda_{gjt} = \frac{\left(P_{jt} \prod_{h=1}^{H} \left(S_{gh}\right)^{T_{hjt}}\right)^{\theta}}{\sum_{j'} \left(P_{j't} \prod_{h=1}^{H} \left(S_{gh}\right)^{T_{hj't}}\right)^{\theta}}. \tag{3}$$

Moreover, the average wages of group g workers equal

$$\overline{W}_{gt} = \Gamma\left(1 - \frac{1}{\theta}\right) \cdot \left(\sum_{j=1}^{J} \left(P_{jt} \prod_{h=1}^{H} \left(S_{gh}\right)^{T_{hjt}}\right)^{\theta}\right)^{1/\theta}, \tag{4}$$

where $\Gamma\left(\cdot\right)$ is the Gamma function.

Given these preferences, we can write the market-clearing condition for occupation j output, equating expenditures on the occupation's output to the wage bill of workers who are employed in occupation j:

$$\underbrace{\frac{\xi_j \cdot (P_{jt})^{1-\rho}}{\sum_{j'=1}^{J} \xi_{j'} (P_{j't})^{1-\rho}}}_{\text{(A)}} \underbrace{\sum_{g=1}^{G} \overline{W}_{gt} L_{gt}}_{\text{(B)}} = \sum_{g=1}^{G} \underbrace{\Gamma\left(1-\frac{1}{\theta}\right)\left(\sum_{j'=1}^{J}\left(P_{j't}\prod_{h=1}^{H}(S_{gh})^{T_{hj't}}\right)^{\theta}\right)^{1/\theta}}_{\text{(C)}} \underbrace{\lambda_{gjt} L_{gt}}_{(D)}. \tag{5}$$

The left-hand side of Equation 5 contains two terms: the share of total expenditures on occupation j output (term A), multiplied by total expenditures (term B). On the right-hand side, the total wage bill of workers in occupation j is computed as the sum of the wage bill of group g workers in occupation j; which in turn equals the product of the average wage of group g workers (term C) and the number of group workers (term D).

We next provide the definition of equilibrium in the model and how we use it to compute changes in the equilibrium

Equilibrium definition In each period t, an equilibrium of our economy is a set of worker allocations $\{\lambda_{gjt}\}_{g,j}$ and prices of occupational output $\{P_j\}$ that satisfy optimal labor supply and occupational output market clearing as determined by Equations 3 and 5.

To evaluate the impact of changes in occupations, we apply the exact hat-algebra approach proposed by Dekle, Eaton, and Kortum (2008). Hence, we our wage equation,

$$\widehat{W}_{gj} = \hat{P}_j \prod_{h=1}^{H} (S_{hi})^{T_{hj}\left(\hat{T}_{hj}-1\right)},$$

our optimal sorting equation,

$$\hat{\lambda}_{gj} = \frac{\left(\widehat{W}_{gj}\right)^{\theta}}{\sum_{j'} \lambda_{gj'} \left(\widehat{W}_{gj}\right)^{\theta}},$$

our average wage equation,

$$\widehat{\overline{W}}_g = \left(\sum_j \left(\widehat{W}_{gj}\right)^{\theta} \lambda_{gj}\right)^{1/\theta},$$

and our market clearing conditions,

$$\sum_{g=1}^{G} \widehat{\overline{W}}_g \Xi_g = \sum_{g=1}^{G} \widehat{\overline{W}}_g \hat{\lambda}_{gj} \chi_{gj},$$

from levels into proportional changes, denoted by hats, between the baseline and the counterfactual.

In this new system of equations, $\left\{\widehat{T}_{hj}\right\}$ are the forcing variables. These are observable in the data, together with the employment shares, λ_{gj}, and the shares in total earnings accrued to group g workers and to group g workers in occupation j, Ξ_g and χ_{gj}. The system determines the counterfactual changes in the employment shares, $\hat{\lambda}_{gj}$, occupational output prices, $\hat{P}_j$, and efficiency wages for group g workers in occupation j, $\widehat{W}_{gj}$.

3.3 Relation to the RIF Regression Approach

Before moving on to our main quantification, we compare our approach to the Recentered-Influence-Function approach which has also been used to think about task contents as drivers of inequality (e.g. in Firpo, Fortin, and Lemieux, 2014). In Appendix A, we outline the assumptions necessary to explicitly link our wage Equation 2 to that used in the RIF approach.[5]

Fortin, Lemieux, and Firpo (2011) introduce a method with which to decompose changes in the distribution of earnings across points in time on the basis of worker and occupational attributes. This method, which can be thought of as an extension of a Oaxaca-Blinder decomposition, breaks down changes over time in any quantile of the earnings distribution into the contribution of observable changes in worker and occupational characteristics (the "composition" effect) and the contribution of implicit changes in the rewards to those characteristics (the "wage structure" effect). One can further break down the composition and wage structure effects into the parts belonging to each of these characteristics (e.g., worker's educational attainment, occupational task content, etc.).

The RIF regression decompositions developed by Firpo, Fortin, and Lemieux (2014) are a flexible accounting device, helpful in determining whether observable characteristics of workers and occupations can account for changes in the earnings distribution. Nevertheless, it has some limitations. One is that the residual in the wage equation contains the contribution to wages of workers' sorting optimally across occupations. A second limitation resides in the interpretation of the decomposition in the face of general equilibrium adjustments. As noted by Firpo, Fortin, and Lemieux (2014), in response to a shock to the task demands, worker mobility across occupations will tend to limit changes in occupation prices, and this approach will attribute wage variation to changes in the returns to skills.

Our model is designed to grapple both with occupational sorting and general equilibrium effects. The model does, however, impose parametric assumptions on workers' idiosyncratic

[5] Through the lens of our model, the coefficients on observable worker attributes in the RIF regressions encapsulate i) the attributes' contribution to task-specific skills and ii) the content of each task in an occupation. On the other hand, in the model task prices' (tasks' marginal contribution to occupational output) do not contribute to inequality once the occupation price is accounted for. The RIF regressions, however, are more flexible, as they accommodate within-occupation variation in task prices.

ability to work in each potential occupation—i.e., on the sources of sorting beyond what we can measure with task contents. We therefore view the two approaches as complementary.

4 Quantification

With this model, we evaluate the impact of change in the demand for tasks on the equilibrium wage distribution. To do so, we first use data from a single period (1960) in combination with Equations 3 and 4 to estimate demographic groups' skills in producing tasks. Then, with our measures of changes in occupations' task content, we compute the counterfactual change in the distribution of wages, between 1960 and 2000, that is due only to changes in the demand for tasks. Appendix B delineates our algorithm to recompute the equilibrium in response to changes in the Ts. Throughout the remainder of the section, we fix $\theta = \rho = 1.78$, following Burstein, Morales, and Vogel (2015).

In estimating our model, we need to map task-related keyword frequencies, as observed in the newspaper data ($\tilde{T}_{hjt}$), to our production function parameters (T_{hjt}). To ensure that our production function parameters sum to one for each occupation-year, we apply the following transformation: $T_{hjt} = \tilde{T}_{hjt} \cdot \left(\sum_{h'} \tilde{T}_{h'jt}\right)^{-1}$.

In this Section we bring the model to the data discussed in Section 2. We proceed in two steps. First, we estimate the key parameters of the model which govern each group g's comparative advantage across tasks and, therefore, across occupations. To do so, we match the model to data in 1960. Second, we feed into this estimated model the observed changes in task content between 1960 and 2000 and ask whether these changes can account for the observed evolution of inequality.

For each task h, we parameterize the relationship between skills and educational and demographic observable variables as follows:

$$\log S_{gh} = a_{h,gender} \cdot D_{gender,g} + a_{h,edu} \cdot D_{edu,g} + a_{h,exp} \cdot D_{exp,g}. \tag{6}$$

Here, $D_{gender,g}$, $D_{edu,g}$, and $D_{exp,g}$ are dummies for gender, education, and experience, which define the categories g. We estimate these a parameters via a method of moments procedure.[6]

Our estimation recovers $a_{h,gender}$, $a_{h,edu}$, and $a_{h,exp}$ to minimize:

$$\sum_{g=1}^{G} \sum_{j=1}^{J} \omega_{gj}^{\lambda} \left(\log \lambda_{gj,1960} - \log \lambda_{gj,1960}^{\text{data}}\right)^2 + \sum_{g=1}^{G} \omega_g^W \left(\log \overline{W}_{g,1960} - \log \overline{W}_{g,1960}^{\text{data}}\right)^2. \tag{7}$$

[6] As in the previous subsection, we have two gender, five educational, and four experience groups, with "male" as the omitted gender category, "Some College" as the omitted educational category, and 20-29 years as the omitted experience category.

In Equation 7, $\lambda_{gj,1960}^{\text{data}}$ and $\log \overline{W}_{g,1960}^{\text{data}}$ refer to the observed occupational shares and average wages.[7] The ω_{gj}^{λ} and ω_{g}^{W} are weights, characterizing the relative importance of each moment in our estimation. We compute these weights as the inverse of the variance of the moment across 40 bootstrapped samples, re-sampling separately from the decennial census (to recover uncertainty on workers' earnings and occupational choices) and from the newspaper text (to recover uncertainty on occupations' task measures).

Table 1 reports our estimates of groups' ability to perform each task. These estimates of comparative advantage are identified primarily from worker groups' sorting across occupations. For example, relative to other groups, workers with higher educational attainment tend to have a comparative advantage in occupations that are relatively intensive in nonroutine analytic tasks; high school graduates and workers with some college tend to have a comparative advantage in occupations intensive in routine cognitive tasks; and high school dropouts have an advantage in occupations intensive in routine and nonroutine manual tasks. These estimates also imply that groups have a comparative advantage over tasks but, since tasks are bundled, ultimately what matters for earnings is the contribution of each task — measured by T_{hj} — to occupational output.[8]

With the aim of substantiating these estimates, we make two points. First, as an assessment in-sample fit of the model, we compare $\log \lambda_{gj,1960}$ and $\log \lambda_{gj,1960}^{\text{data}}$ (across the 3240 group-occupation observations), and $\log \overline{W}_{g,1960}$ and $\log \overline{W}_{g,1960}^{\text{data}}$ (across the 40 groups). The correlation between the model-estimated occupational shares and the observed shares is 0.65, while the correlation between $\log \overline{W}_{g,1960}$ and $\log \overline{W}_{g,1960}^{\text{data}}$, is 0.98. Second, identification of the parameters in Table 1 follows transparently from workers' sorting patterns. Based on our set-up, the share of workers will be high for occupations in which their skills

[7] Our estimation relies on data from 3240 (=$40 \cdot 81$) moments—representing information from 40 groups (2 genders, 5 educational categories, and 4 experience categories) and 81 occupations—to identify 121 parameters: 81 occupation-fixed effects plus 40 a coefficients.

[8] More specifically, focusing on the observable component of comparative advantage, group g will have a comparative advantage in occupation j, relative to group g' and occupation j', if

$$\sum_{h=1}^{H} (T_{hj} - T_{hj'}) (\log S_{gh} - \log S_{g'h}) > 0.$$

For instance, consider comparing groups' comparative advantage across three occupations: Physical Scientists (SOC=1920, an occupation intensive in nonroutine analytic tasks), Financial Clerks (SOC=4330, an occupation intensive in routine cognitive tasks), and Material Movers (SOC=5370, intensive in routine manual and nonroutine manual tasks). The T vectors associated with these three occupations are (with tasks ordered alphabetically): $(0.73, 0.18, 0.03, 0.04, 0.02)$, $(0.14, 0.29, 0.02, 0.52, 0.04)$, and $(0.25, 0.49, 0.10, 0.07, 0.09)$. Based on these task measures in conjunction with the formula above, we would conclude that college graduates have a comparative advantage relative to all other education groups in the Physical Scientist occupation, high school graduates have a comparative advantage in the Financial Clerks occupation, and workers with less than high school education have a comparative advantage in the Material Movers occupation.

Table 1: Estimates of Skills

	Nonroutine Analytic	Nonroutine Interactive	Nonroutine Manual	Routine Cognitive	Routine Manual
Gender					
Female	-1.708	-0.299	-0.658	3.209	-3.800
	(0.117)	(0.083)	(0.595)	(0.210)	(0.850)
Education					
$<$ HS	-2.159	-0.494	4.194	-1.558	3.884
	(0.158)	(0.070)	(0.431)	(0.132)	(0.543)
High School	-1.092	-0.368	2.413	0.322	1.943
	(0.057)	(0.034)	(0.229)	(0.044)	(0.348)
College	1.449	0.372	-3.172	-1.369	0.750
	(0.103)	(0.060)	(0.939)	(0.141)	(1.432)
Post-Graduate	1.285	0.290	-4.274	-1.504	5.006
	(0.174)	(0.117)	(1.556)	(0.445)	(1.663)
Experience					
0-9 Years	-0.299	-0.859	1.312	0.714	-3.078
	(0.066)	(0.034)	(0.243)	(0.074)	(0.284)
10-19 Years	-0.006	-0.209	0.409	0.273	-0.800
	(0.022)	(0.014)	(0.072)	(0.021)	(0.087)
30+ Years	-0.076	0.045	0.416	0.239	-0.851
	(0.032)	(0.019)	(0.101)	(0.032)	(0.134)

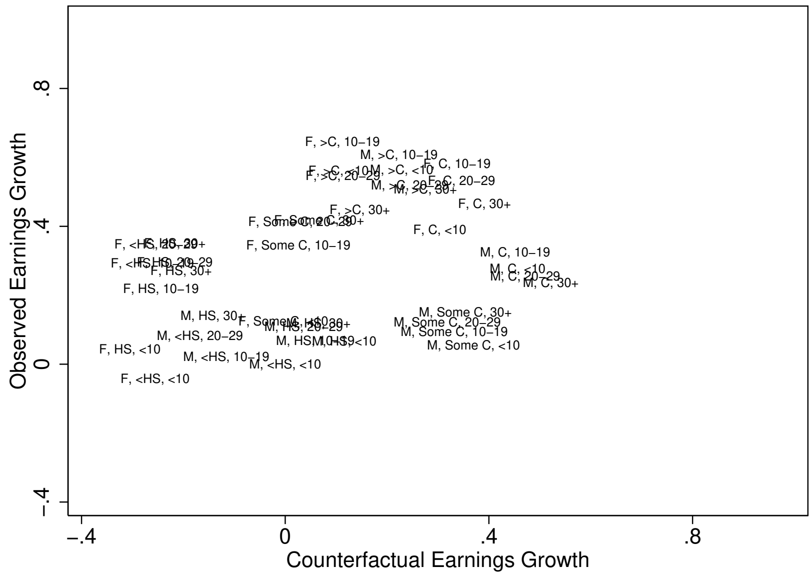
Notes: To compute the standard errors, we re-sampled 40 times from the 1960 decennial census to recover sampling uncertainty on earnings and occupational choices, and re-sampled 40 times from our newspaper text to recover sampling uncertainty on occupations' task measures. For each bootstrapped sample, we recomputed the empirical occupational shares and group wages and then found the combination of parameters that minimized Equation 7. The omitted demographic groups are males, workers with Some College, and workers with 20-29 years of potential experience.

will be used intensively. For example, mirroring the estimates in Table 1, the correlation between the average number of words related to routine cognitive tasks in occupation j and an occupation-group's log $\lambda_{gj,1960}$ is substantially higher for women compared to men (0.32 versus -0.03); for high school graduates compared to post-graduates (0.24 versus -0.01); and for workers with fewer than 10 years of experience compared to workers with 30+ years of experience (0.15 versus 0.06).

Counterfactual Earnings Distribution

Having estimated the distribution of skills, we now compute the counterfactual dispersion in wages that would otherwise have been obtained in succeeding decades—in 1970, 1980, 1990, and 2000—as a result of changes in the demand for tasks over our sample period. To do so, we consider an exogenous change in the T_{hjt} production function parameters (which, again,

Figure 1: Counterfactual and Observed Earnings Growth, 1960-2000, By Demographic Group



Notes: Each point gives the growth in wages for one of the 40 g groups The first character—"M" or "F"—describes the gender; the second set of characters—"<HS," "HS," "SC," "C," or ">C"—the educational attainment; and the third set of characters the number of years — "0" for 0-9, "1" for 10-19, "2" for 20-29, "3+" for ≥30 — of potential experience for the demographic group.

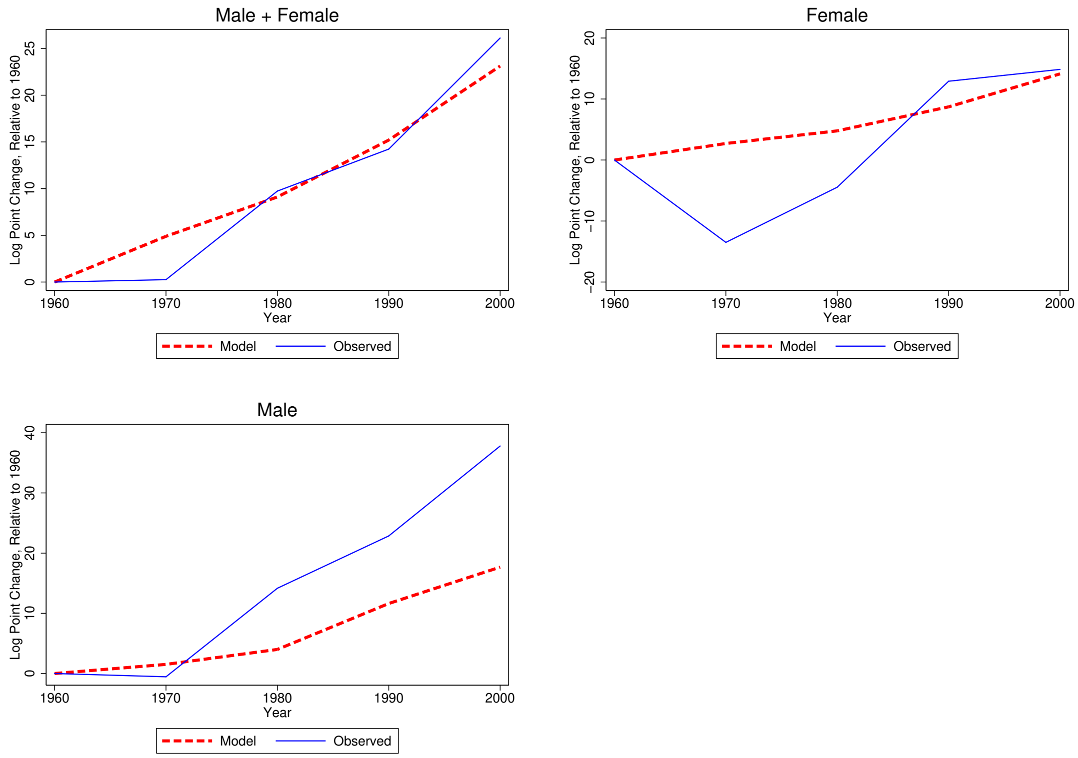
we measure through the changes in task-related keyword frequencies). Stemming from these changes in the demand for tasks, we compute the counterfactual equilibrium that would have obtained, fixing $\log S_{gh}$ to the values we estimated in the previous subsection, but allowing T_{hjt} to change over time.[9]

Figure 1 plots the growth in $\overline{W}_{gt}$ between 1960 and 2000, both as observed in the data and according to our counterfactual exercise. Wages increase most quickly for workers with college degrees or with post-graduate education. According to our estimated model, wage growth is fastest for high-education demographic groups because these groups have an advantage in nonroutine analytic and interactive tasks as of 1960, and the relative demand for these tasks has increased over the subsequent 40 years.

Increases in between-group inequality signify larger overall inequality. Figure 2 plots changes in 90-10 inequality, both the observed growth rate and the counterfactual growth rates that are due only to changes in T_{hjt}. Between 1960 and 2000, 90-10 inequality increased by 26 log points for the entire population, 38 log points for male workers, and 15 log points

[9] In the exercises in this section, we also fix the sizes of the 40 demographic groups to their 1960 values. An alternate procedure would allow these labor supplies to vary throughout the sample according to observed demographic changes. This alternate procedure yields similarly large increases in 90-10 inequality in the counterfactual equilibrium, compared to those reported in Figure 2.

Figure 2: Counterfactual and Observed 90-10 Inequality, 1960-2000



Notes: Each panel plots the change in 90-10 earnings inequality, as observed in the decennial census (solid line), and are due only to changes in the demand for tasks (dashed line).

for female workers. According to our counterfactual exercises, within occupation changes in task demand account for much of the observed increase in 90-10 inequality: 23 log points for the entire sample, 15 log points for female workers, and 16 log points for male workers.

Discussion

Occupational choice and wage data from the beginning of our sample suggest that workers with certain demographic characteristics (e.g., workers with a college degree) have an advantage in the occupations rich in tasks which have happened to grow over the subsequent forty years. Since these demographic groups were already highly remunerated in 1960, changes in the demand for tasks have increased earnings inequality.

Where do these shifts in the demand for tasks come from? Our model omits capital as an input in the production of tasks. Likely, some of the differential growth rates of T_{hjt} across task groups reflect, at a more fundamental level, changes in capital prices. To the extent that the elasticity of substitution between workers and capital is relatively high in the production of routine tasks, a decline in the price of capital will be equivalent to a reduction in the demand for worker-performed routine tasks.

Insofar as employers' innovation and technology adoption decisions are a function of the supply of workers of different skill types, changes in the T_{hjt} not only embody exogenous changes in the demand for tasks (which has been our interpretation here), but also reflect changes in the skill composition of labor supply.[10] An interesting topic for future work would be to build on our model of occupations as a bundle of tasks, endogenizing changes in the T_{hjt} while acknowledging that workers and capital combine to perform each individual task and that innovations which lead to these declines in capital prices are potentially a response to the changing composition of the workforce.

In the context of these results, we can also revisit the distinction relative to the RIF-regression approach. In our equilibrium quantification, worker sorting plays a key role. Rewards to skills are occupation specific (and determined by the occupations' associated task content), and sorting amplifies the effect of observed changes in tasks on inequality.[11] Moreover, while heavily parameterized, the model takes into account the labor responses to changes in task prices induced by the shifts to the task contents, T_{hjt}. These general equilibrium effects also tend to generate more inequality explained by tasks in the model relative to the statistical decomposition.

On the other hand, we only evaluate the effect of within-occupation changes in task demands in the model, while other shocks to fundamentals could also shape the income distribution. These other shocks include changes in the demand for occupations (measured by ξ_j in the model), and potentially represent a source of widening income inequality that is not captured in our model. Moreover, to the extent that the price of individual tasks vary within occupations, the RIF regressions will capture a source of inequality that the model cannot.

5 Conclusion

In this paper, we demonstrate that within-occupation task changes are fundamentally important in an examination of the evolution of income inequality. We construct and estimate Roy model in which bundled tasks determine workers' comparative advantage. The model allows us to quantify the impact of changes in task contents within occupations—a margin that has not been examined before. We find that these changes alone account for a 23 log

[10] In Atalay, Phongthiengtham, Sotelo, and Tannenbaum (2018), we explore in detail how the arrival of new technologies has transformed individual jobs and the labor market as a whole, as well as how these changes have impacted welfare and inquality.

[11] In principle, sorting on unobservable characteristics could lead the statistical decomposition to either overstate or understate the role of tasks in accounting for increased inequality. If, over time, there is increased sorting of workers with high unobserved ability into jobs at the top of the distribution, then the decomposition's wage structure effects will under-represent the contribution of task price changes to inequality. Firpo, Fortin, and Lemieux (2014) argue that this is likely the case. See their Section I.B.

point increase in labor income inequality (as measured by the 90-10 ratio) between 1960 and 2000.

References

Acemoglu, D., and D. Autor (2011): "Skills, Tasks and Technologies: Implications for Employment and Earnings," *Handbook of Labor Economics*, 4, 1043–1171.

Atalay, E., P. Phongthiengtham, S. Sotelo, and D. Tannenbaum (2018): "New Technologies and the Labor Market," vol. 97 of *Journal of Monetary Economics*. Carnegie-Rochester-NYU Conference Series, forthcoming.

——— (Forthcoming): "The Evolution of Work in the United States," *American Economic Journal: Applied Economics*.

Autor, D. H., and D. Dorn (2013): "The Growth of Low-Skill Service Jobs and the Polarization of the US Labor Market," *American Economic Review*, 103(5), 1553–97.

Autor, D. H., L. Katz, and M. Kearney (2005): "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.

Burstein, A., E. Morales, and J. Vogel (2015): "Accounting for Changes in Between-Group Inequality," Working Paper 20855, National Bureau of Economic Research.

Dekle, R., J. Eaton, and S. Kortum (2008): "Global Rebalancing with Gravity: Measuring the Burden of Adjustment," *IMF Staff Papers*, 55(3), 511–540.

Deming, D. (2017): "The Growing Importance of Social Skills in the Labor Market," *Quarterly Journal of Economics*, 132(4), 1593–1640.

Firpo, S., N. M. Fortin, and T. Lemieux (2014): "Occupational Tasks and Changes in the Wage Structure," IZA Discussion Papers 5542, Institute for the Study of Labor (IZA).

Fortin, N., T. Lemieux, and S. Firpo (2011): "Decomposition Methods in Economics," *Handbook of Labor Economics*, 4, 1–102.

Galle, S., A. Rodriguez-Clare, and M. Yi (2017): "Slicing the Pie: Quantifying the Aggregate and Distributional Effects of Trade," Discussion paper, National Bureau of Economic Research.

Heckman, J. J., and J. Scheinkman (1987): "The Importance of Bundling in a Gorman-Lancaster Model of Earnings," *Review of Economic Studies*, 54(2), 243–255.

Heckman, J. J., and G. Sedlacek (1985): "Heterogeneity, Aggregation, and Market Wage Functions: An Empirical Model of Self-Selection in the Labor Market," *Journal of Political Economy*, 93(6), 1077–1125.

Lagakos, D., and M. E. Waugh (2013): "Selection, Agriculture, and Cross-Country Productivity Differences," *American Economic Review*, 103(2), 948–980.

Michaels, G., F. Rauch, and S. J. Redding (2016): "Tasks and Technology in the United States 1880-2000," Discussion paper, National Bureau of Economic Research.

Ruggles, S., K. Genadek, R. Goeken, J. Grover, and M. Sobek (2015): "Integrated Public Use Microdata Series: Version 6.0," Minneapolis, MN: Historical Census Projects, University of Minnesota.

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.

Yamaguchi, S. (2012): "Tasks and Heterogeneous Human Capital," *Journal of Labor Economics*, 30(1), 1–53.

Online Appendix

In this section, we describe the assumptions necessary to derive the empirical specification for our RIF regression decomposition from Equation 2. We then delineate two algorithms: In the first, we compute the counterfactual changes in groups' wages, and in the second we apply those changes in groups' wages to compute changes in the overall distribution (within-group and between group) of wages. Throughout this section use $\hat{X}$ to refer to the ratio of variable X from one period to the next: $\hat{X} = \frac{X_{t+\tau}}{X_t}$.

A Linking the Model of Occupations as a Bundle of Tasks to the RIF Regressions

The goal of this appendix is to derive conditions under which there is a clear map between our approach and the decompositions based on RIF regressions. We begin by relating the price of occupation-j specific output with the prices of its constituent tasks:[12]

$$\log P_{jt} = \log \pi_{0jt} + \sum_{h=1}^{H} T_{hjt} \log \pi_{hjt}. \tag{8}$$

In our model, perfect competition and homogeneous occupational output ensure that there is a single price for each occupation. Thus, although task prices π_{hjt} may vary by group reflecting the marginal value of each task in a production unit, these prices do not contribute to inequality once one takes into account the output price P_j. In other words, task prices only generate inequality between occupations. In contrast, the RIF regression specification, which allows for different task prices at different quantiles of the income distribution. One way to justify this, more flexible specification, is to assume that workers at different quantiles work in different occupations. Another is to consider a more flexible model where the unobserved shocks are not Hicks neutral in a given occupation, but rather task-specific.

Bearing in mind that additional flexibility, we can use the pricing Equation (8) in the wage Equation (2) to show how wages relate to tasks, task prices, and skills:

$$\log W_{ijt} = \log \pi_{0jt} + \sum_{h=1}^{H} T_{hjt} \log \pi_{hjt} + \sum_{h=1}^{H} T_{hjt} \log S_{gh} + \log \epsilon_{ijt}. \tag{9}$$

[12] This is akin to the formulation in Yamaguchi (2012). To motivate this equation, suppose that workers cannot "unbundle" their tasks in the sense of Heckman and Scheinkman (1987). To the extent that tasks cannot be "unbundled," task prices will differ across occupations.

We next assume that individual i's ability to perform task h at time t is linearly related to a set K of observable characteristics so that the underlying skill of group g individuals in performing task h can be written as $\log S_{gh} = \sum_{k=1}^{K} b_{hk}\tilde{S}_{gk}$. With this assumption, Equation (9) implies:

$$\log W_{ijt} = \log \pi_{0jt} + \sum_{h=1}^{H} T_{hjt} \log \pi_{hjt} + \sum_{k=1}^{K} \alpha_{kjt}\tilde{S}_{kg} + \log \epsilon_{ijt}, \tag{10}$$

where, again, $\tilde{S}_{gk}$ is an observable skill characteristic k for an individual in group g.

The flexible wage equation at the basis of the RIF approach would be equivalent to Equation 10, except the latter allows for prices to vary across occupations j. However, insofar as workers in different occupations are concentrated at different points in the wage distribution, the coefficient estimates in a RIF regression will vary at different quantiles.

B Algorithms

B.1 Algorithm to Compute Counterfactual Changes in Wages

1. Start with an occupational price guess $\hat{P}^{[0]}$.

2. Compute

$$\hat{W}_{gj}^{[1]} = \hat{P}_j^{[0]} \prod_{h=1}^{H} \left(S_{hi}\right)^{T_{hj}\left(\hat{T}_{hj}-1\right)}.$$

$$\hat{\lambda}_{gj}^{[1]} = \frac{\left(\hat{W}_{gj}^{[1]}\right)^{\theta}}{\sum_{j'} \lambda_{gj'} \left(\hat{W}_{gj}^{[1]}\right)^{\theta}}.$$

$$\widehat{\bar{W}}_g^{[1]} = \left(\sum_j \left(\hat{W}_{gj}^{[1]}\right)^{\theta} \lambda_{gj}\right)^{1/\theta}.$$

3. Compute the excess demand function

$$\begin{aligned} Z_j^{[1]} &= \hat{E}^{[1]} - \sum_{g=1}^{G} \widehat{\bar{W}}_g^{[1]} \hat{\lambda}_{gj}^{[1]} \chi_{gj} \\ &= \sum_{g=1}^{G} \widehat{\bar{W}}_g^{[1]} \varXi_g - \sum_{g=1}^{G} \widehat{\bar{W}}_g^{[1]} \hat{\lambda}_{gj}^{[1]} \chi_{gj}, \end{aligned}$$

   where $\varXi_g$ and χ_{gj} are, respectively, the initial-equilibrium share of income earned by group g individuals and the share of group g occupation j in total labor payments.

4. Compute the new prices

$$\hat{P}_j^{[1]} = \hat{P}_j^{[0]} + \nu Z_j^{[1]}$$

for some small adjustment number ν. We also need to normalize $P_1^{[1]} = 1$.

5. Check for convergence. If there is no convergence, go back to 2.

B.2 Algorithm to Simulate the Distribution of Wages

Given $\bar{W}_{g,1960}$ and $\widehat{\bar{W}}_g$ (constructed from the previous algorithm), we now know $\bar{W}_{g,t}$. Note that the distribution of wages for each type g in this model is Frechet, with parameters $\left(\bar{W}_{g,t}, \theta\right)$. To get at the distribution of wages within the model:

1. Fix a large number $\bar{L}^{sim}$ that controls the size of the simulation. We use $\bar{L}^{sim} = 10^6$.

2. For each group g, sample $\bar{L}^{sim} \times \sum_{j=1}^{J} \lambda_{gj}$ wages. To do so, draw for each observation in group i a unit exponential $\mathcal{U}$ and then apply the transformation $W = \bar{W}_{g,t} \cdot \mathcal{U}^{-1/\theta}$.

3. The vector of all W is a sample of wages in this economy.