BLS WORKING PAPERS

U.S. Department of Labor
U.S. Bureau of Labor Statistics
Office of Employment Research and Program Development

Outsourcing, Occupationally Homogeneous Employers, and Growing Wage Inequality in the United States

Elizabeth Weber Handwerker, U.S. Bureau of Labor Statistics

Working Paper 522
October 2021

JEL codes: J31, J53, M51, L24, M55
Keywords: outsourcing, employer homogeneity, wage inequality, contracting out, employment polarization

Outsourcing, Occupationally Homogeneous Employers, and Growing Wage Inequality in the United States

Elizabeth Weber Handwerker
U.S. Bureau of Labor Statistics

October, 2021

Abstract: This paper develops measures of the occupational homogeneity of employers as indicators of outsourcing. Findings are threefold. First, workers—particularly those in low-wage occupations who worked in smaller establishments—saw their employing establishments become more occupationally homogeneous during 2004-2019. Second, wages are strongly related to occupational homogeneity, particularly for workers in low-wage occupations. Third, changes in the occupational homogeneity of workplaces are an important contributor to growing wage inequality among workers over the first half of this period. The growing separation of workers in low-wage occupations into different employers from workers in high-wage occupations is an important part of wage inequality growth.

I am grateful to seminar participants at the American Economic Association 2014 Annual Meetings, the University of Maryland, the Economic Policy Institute, the Federal Reserve Board of Governors, the Society of Labor Economists 2016 and 2017 Annual Meetings, the Center for Equitable Growth, the NBER 2018 Summer Institute, the BRIQ Institute, and the Bureau of Labor Statistics for comments, with special thanks to Matt Dey, Katharine Abraham, Andy Garin, David Autor, Erica Groshen, Sue Houseman, David Levine, Anne Polivka, Alex Bryson, David Piccone, David Weil, Nicole Fortin, Fernando Rios-Avila, and anonymous referees for particularly helpful comments and suggestions. This paper grew out of a project begun jointly with James Spletzer, and I thank him for his many contributions. Any opinions and conclusions herein are those of the author and do not necessarily reflect the views or policies of the U.S. Bureau of Labor Statistics. Research assistance from Lowell Mason and David Piccone is gratefully acknowledged.

I. Introduction

Growing inequality of wages, particularly between employers, has been a key feature of the labor market in recent decades. Many changes in the labor market have been examined as potential sources of this inequality growth—including the decline of manufacturing, the role of technology in replacing employer demand for routine work, and the increased potential for imported goods and services to replace domestic production. This paper examines an additional source of growing wage inequality: the changing distribution of occupations between employers as the organization of production changes, with employers retaining certain types of work within the workplace and outsourcing other work.

Much evidence shows that establishments play an important role in determining individual wages, beyond the role of individual workers' characteristics (Groshen 1991a, 1991b; Bronars and Famulari 1997; Abowd, Kramarz, and Margolis 1999; Lane, Salmon, and Spletzer 2007; Card, Heining, and Kline 2013). Several authors have used employer microdata to study growing variability in earnings in the U.S. from the mid-1970s to the early 2000s, and have found it due more to variation between establishments than to variation within establishments (Davis and Haltiwanger 1991; Dunne, Foster, Haltiwanger, and Troske 2004; Barth, Bryson, Davis, and Freeman 2016; Handwerker and Spletzer 2016; and Song, Price, Guvenen, Bloom, and von Wachter 2019),[1] while the increased sorting of high-paid workers to high-paying employers drives much of the growth in pay inequality between employers (Song, Price, Guvenen, Bloom, and von Wachter, 2019). The results in this paper show that occupational

[1] There is a large and growing literature on wage inequality growth in Europe, based on employee-employer linked data, including Card, Heining, and Kline (2013), who emphasize the role of increased worker sorting between employers in explaining wage inequality growth in Germany.

homogeneity—a specific form of worker sorting—is a key explanation for this growth in between employer wage inequality. More and more workers in low-wage occupations are employed in different workplaces from workers in other occupations, exacerbating differences in their pay.

The intersection of growing underlying wage inequality and the business environment in the United States can make it profitable for employers to focus on employing either low or high wage workers. Growing wage inequality among workers has arisen from such sources as the changing composition of the workforce and changing returns to education and experience,[2] the growing inequality within education and skill groups[3], and the differential impact of technology on the worker skill distribution[4]. As wages for different kinds of work become less equal, employers face regulations requiring nondiscrimination across employees in the coverage of pension, health insurance and other benefits (EBRI 2009, Perun 2010),[5] increasing incentives to contract out work that pays very different wages from the work of other employees. Moreover, social norms may make it more acceptable for employers to contract out work rather than pay very different wages to employees doing different kinds of work (Weil 2014).

Other potential reasons for businesses to outsource work include increasing ability to smooth workload, economies of scale available to providers of specialized services (Abraham and Taylor, 1996), or a focus on "core competencies" enabled by technologies for specifying and

[2] Bound and Johnson 1992, Katz and Murphy 1992, Lemieux 2006
[3] Juhn, Murphy, and Pierce 1993, Katz and Autor 1999
[4] Juhn, Murphy, and Pierce 1993, Acemoglu 2002, Autor, Katz, and Kearney 2006, 2008
[5] Perun (2010) lists a variety of employment benefits which receive favorable tax treatment and are required to be available to low-wage as well as high-wage employees of each employer.

monitoring work done by outsiders (Weil 2014). However, to the extent that labor cost savings and avoidance of efficiency wages or rents for occupations with low wages in the labor market are key reasons for outsourcing, it can lead to employers specializing in high or low-wage work, and result in growing wage inequality between establishments. Goldschmidt and Schmeider (2017) show labor cost savings to be a primary reason for outsourcing in Germany, as outsourced workers lose firm-specific rents, while Drenik, Jäger, Plotkin, and Schoefer (2021) study this same phenomenon for the outsourcing of work to temp agencies in Argentina, and Bilal and Lhuillier (2021) model its impact in France. In three well-defined occupational categories, Goldschmidt and Schmeider find that losses of such firm-specific rents can account for 9% of all growth in German wage inequality from 1985 to 2008.

In U.S. data, direct measures of outsourcing are not generally available. Researchers have instead focused on particular industries or occupations associated with performing support tasks for other businesses. Dey, Houseman, and Polivka (2010) show a marked increase in various measures of outsourcing in recent years such as trends in temporary help or employment services. Estimates from several sources show these industries roughly doubling in size from 1992 to 2002. They also document an increase in the employment share of occupations associated with outsourced labor services, such as school bus and truck drivers in the transportation industry and accountants in the business services industry. Yet these measures only capture a fraction of outsourcing—that which occurs in these specific industries. Dube and Kaplan (2010) use individual-level data to show the impact of outsourcing on wages and benefits for janitors and guards, but again, their measures can only capture outsourcing of a narrow set of occupations.

This paper develops economy-wide measures of outsourcing in the United States, using the homogeneity of occupations by employer, as measured in the detailed microdata of the Occupational Employment and Wage Statistics Survey conducted by the Bureau of Labor Statistics. These measures distinguish between two types of outsourcing, which may have differing impacts on wage inequality. When businesses outsource work to avoid monitoring, hiring, or other costs for occupations in which they have less expertise, there will be less variety overall in the occupations they employ. However, when businesses outsource work to narrow the wage distribution of their employees, the variance of wages predicted from the particular set of occupations they employ will decrease. The impact of the changing distributions of occupations and of employer occupational homogeneity are compared with the effects of other changes in employer characteristics (industry, size, and location) on the overall distribution of wages.

There are three major findings. First, from 2004 through 2019, the occupational homogeneity of employers increased overall, and especially for workers in typically low-wage occupations, after controlling for other employer characteristics. Second, wages are related to the occupational homogeneity of establishments. Workers in more occupationally homogeneous establishments earn lower wages. This relationship holds even after controlling for workers' own occupations and observable characteristics of their employers, and is strongest for workers in occupations typically paid low wages. Third, changes in the distribution of this occupational homogeneity are related to the growth in private-sector wage inequality observed in the data during the first twelve years of this time period. A substantial amount of the growth in ln(wage) variance, as measured in the OEWS data, can be attributed to the growing occupational homogeneity of establishments over this period. Both measures of employer homogeneity—one based on the distribution of occupations by wage levels, and the other a more functional measure of employer homogeneity that ignores wage differences among occupations—matter for growing wage inequality.

The paper is organized as follows: Section II describes measures of occupational homogeneity. Section III describes trends in measured occupational homogeneity of employers. Section IV describes relationships between employer occupational homogeneity and employee wages. Section V describes the impact of the changing distributions of occupation and the occupational homogeneity of employers on wage inequality over time. Section VI concludes.

II. Measuring the Occupational Homogeneity of Employers

I use the term "occupational homogeneity"[6] to describe the variety of occupations employed at a place of business, separate from the tasks performed by individual employees (their occupations), the type of work done at the business (its industry) or the size of the business. Much scholarship on outsourcing (for example Dey, Houseman, and Polivka, 2010; and Erickcek, Houseman, and Kalleberg, 2003) examines particular occupations and particular industries. In contrast, occupational homogeneity is intended as a measure of the variation in work done in all businesses, through the full range of industries in the economy. This section

[6] Earlier versions of this paper referred to the same concept as "occupational concentration."

defines two measures of occupational homogeneity and presents evidence showing that these measures are related to examples in the outsourcing literature.

The two measures of the occupational homogeneity of establishments are very different: (1) a measure involving the overall distribution of occupations, regardless of whether they are high or low paid, and (2) a measure that explicitly models the variation in wages of establishments due to the distribution of occupations employed.

The first measure of occupational homogeneity for establishment *j* at time *t* is constructed with a Herfindahl-Hirschman index of employment, *n*, in each occupation *k* within that establishment, normalized for the overall size of the establishment, N:

(1) $$H_{jt} = \sum_{k=1}^{100} \left(\frac{n_{kjt}}{n_{jt}}\right)^2 \qquad Normalized\ H_{jt} = \frac{(H_{jt} - 1/N)}{(1 - 1/N)}, \text{ or 0 if N = 1}$$

This index uses the 100 minor occupational categories at the 3-digit level of the Standard Occupational Classification system.[7] It varies from 0 (equal representation of all occupations) to 1 (complete homogeneity). Increased occupational homogeneity at the establishment level by this measure indicates that employers are becoming more specialized, consistent with outsourcing work to other employers. Trends in this measure indicate whether establishments throughout the U.S. economy are becoming more homogeneous in the occupations they employ.

[7] Handwerker and Spletzer (2016) studied this type of general occupational homogeneity with Herfindahl-Hirschman indices, using both the detailed 6-digit occupations of the Standard Occupational Classification System (829 categories) and the 2-digit major occupational categories of the Standard Occupational Classification System (22 categories), and found very similar time trends and relationships between occupational classification and wages with broad and detailed versions of this measure.

However, this measure cannot distinguish between specializing in a few occupations typically paid very different wages, such as 29-1000 (Healthcare Diagnosing or Treating Practitioners) and 31-1100 (Home Health and Personal Care Aides; and Nursing Assistants, Orderlies, and Psychiatric Aides), or specializing in a similar number of occupations that are typically paid more similar wages.

In contrast, the second measure of occupational homogeneity is explicitly constructed to capture the similarity or dissimilarity of typical wages for the occupations employed at an establishment. It is the part of the variance of wages for each establishment that would be predicted from the establishment's distribution of employment by occupation, without using information on the actual wages paid at the establishment. Using average log wages for each minor occupational category in each time period, the log wage paid by employer *j* to worker *i* in occupation *k* at time *t* is estimated as $\widehat{w_{ijt}} = \overline{w_{kt}} + \varepsilon_{ijt}$, where $\overline{w_{kt}}$ is the mean log wage for all employees in occupation *k* at time *t* and ε_{ijt} is distributed normally, with mean 0 and standard deviation σ_k. From the occupational distribution of employer j at time t, the estimated mean log wage for j at t is estimated $\widehat{\overline{w_{jt}}} = \frac{\sum_k \sum_{i \in k} \overline{w_{kt}}}{n_{jt}}$, where n_{jt} is the total employment for employer *j* at time *t*, and $i \in k$ denotes observations in which individual *i* has occupation *k*. Again, using only the distribution of occupations employed and the average wages of these particular occupations across all employers at time t, the predicted log wage variance for employer *j* at time *t* is

$$\text{(2)} \qquad \widehat{V_{jt}} = \frac{\sum_i \left(\widehat{w_{ijt}} - \widehat{\overline{w_{jt}}}\right)^2}{n_{jt}} = \frac{\sum_k n_{jkt}\left[\left(\overline{w_{kt}} - \widehat{\overline{w_{jt}}}\right)^2\right]}{n_{jt}} + \frac{\sum_k n_{jkt}\sigma_{kt}^2}{n_{jt}}.$$

This has two terms: the variation in average wages between occupations, and the average of within-occupation log wage variances. The first term in equation (2) is the second measure of occupational homogeneity.

Both of these measures are estimated with the microdata of the Occupational Employment and Wage Statistics (OEWS) Survey for the private sector in the United States for 2004 through 2019. These microdata record the number of employees by wage interval within detailed occupation categories for hundreds of thousands of establishments per year. The OEWS survey is designed to produce estimates of employment and wages in the United States for each detailed occupation, by geography and industry. It covers all establishments in the United States except for those in agriculture, private households, and unincorporated self-employed workers without employees. It is the only survey of its size and scope.

The OEWS collects data for a sample of about 200,000 establishments each November and each May. Sampled establishments are asked to report the number of employees in each occupation by wage interval. As described in Dey and Handwerker (2016), the OEWS uses a complex sample design intended to minimize the variance of published wage estimates for each occupation within industries and geographic areas. Establishments expected to employ rarer occupations or occupations with greater variation in wages have relatively larger probabilities of selection.

In using OEWS data to study wage inequality, it is important to understand that the OEWS data *cannot* measure inequality in the topmost percentiles of the wage distribution.

Wages are reported to the OEWS in intervals. The OEWS program uses the mean of each wage interval for each minor occupational group for each reference period from the National Compensation Survey (NCS) to assign wages for employees in each wage interval. Earnings of individuals at the very top of the wage distribution are topcoded in the OEWS—the uppermost interval in the recent OEWS surveys is "$208,000 and over." Averaged across all years, the uppermost interval contains roughly 1.3 percent of employment. Handwerker and Spletzer (2014) compare wage inequality levels and trends in these OEWS microdata with the wage inequality level and trends in the outgoing rotation group microdata of the CPS, which has been used in many of the most cited studies of wage inequality. They show the interval nature of wage collection in the OEWS has almost no impact on overall wage variance trends. Both this study and the update in Dey, Handwerker, Piccone, and Voorheis (2022) show that OEWS data broadly replicate CPS wage distribution levels and trends: overall wage variances in each year are similar in the reweighted OEWS and CPS microdata until 2016. However, from 2016 to 2019, Dey, Handwerker, Piccone, and Voorheis show there was a more substantial wage variance decline in OEWS data than in CPS data.

The OEWS sample design uses 3 years, or 6 panels of data collection, to produce detailed published estimates of employment and wages, with employment weights benchmarked to employment at the time of the last panel and adjustments to wages based on the BLS Employment Cost Index so that wages refer to wage levels in that last panel. It is not designed to produce time series estimates of either employment or wages for any individual occupation, in part because of changes over time in occupational definitions. This paper uses OEWS microdata from November 2004 (collected from 2001 through 2004), November 2007 (collected from 2005 through 2007), November 2010 (collected from 2008 through 2010), November 2013 (collected from 2011 through 2013), November 2016 (collected from 2014-2016), and November 2019 (collected from 2017-2019). Various adjustments are made to occupations and industries to make them as consistent as possible throughout the period.

Establishments are the sampling units of the OEWS, and so this paper focuses on measures of occupational homogeneity at the establishment level. However, all the main results in this paper have been repeated with measures constructed at the Employer Tax-ID level (EIN), and results are shown in Appendix C.

The Data Appendix contains summary statistics, including the composition of occupations and industries. The average worker has an inflation-adjusted wage of $16.54/hour (in $2000), or a ln(wage) of 2.59, and is observed in an establishment with a measured ln(wage) variance of 0.166. The average normalized Hirfindahl-Hirschman index for workers' establishments is 0.360, and the average predicted variance of ln(wages) estimated from its workers' occupational composition is 0.263. It is unsurprising that the predicted ln(wage) variance based only on the occupations employed at the establishment is higher than the measured ln(wage) variance because of the large literature describing the impact of employer-specific factors on wages. The average part of this predicted variance due to variation in wages between occupations (rather than within occupations) is 0.103.

Table 1 compares the two measures of establishment-level occupational homogeneity for several occupation-industry groups studied as examples of outsourcing by Abraham and Taylor

(1996); Dube and Kaplan (2010); Dey, Houseman, and Polivka (2010); Weil (2014); and Goldschmidt and Schmeider (2015): the entire food preparation and serving major occupational group, janitors, security guards, truck drivers, accountants, computer occupations, engineering occupations, and lawyers. Outsourcing of workers in these occupations means that they are employed in the specialty industries of food services, janitorial services, security guard services, truck transportation, accounting services, computer services, engineering services, or law offices, rather than the industry of the business to which they provide these services. Table 1 shows that for every single one of these example occupations or occupation groups, the normalized Herfindahl-Hirschman indices for employers of these workers (as defined in equation (1)) are higher, on average, indicating greater occupational homogeneity of employers, when they are employed in their specialty industry than when they are employed in other industries. Moreover, for every example occupation except lawyers (the smallest and highest paid of these examples), the partial predicted variances of wages based on the occupational distribution of their employers are lower, on average, indicating greater occupational homogeneity of employers, when they are employed in their specialty industry than when they are employed in other industries. Both measures of occupational homogeneity measures defined in this section—designed to measure outsourcing across all occupations and industries—indicate greater occupational homogeneity in the relevant industries to which workers are outsourced for the specific occupations studied in the outsourcing case-study literature.

III: Trends in Occupational Homogeneity Measures

Understanding trends in occupational homogeneity measures is complicated by contemporaneous changes in the overall occupational composition of the labor force. As described by Autor, Katz, and Kearney (2006, 2008), among others, employment in typically low-wage and typically high-wage occupations has increased, while employment in many typically middle wage-occupations has decreased. Figure 1 shows employment over time for occupational quintiles in the OEWS. Employment polarization is clear in the OEWS data: there is an increasing fraction of employment over time in the top quintile, with a decreasing fraction of employment in the middle three quintiles. This polarization means that if we entirely ignore the grouping of employment into establishments and if occupation-level wages stay constant, the portion of the variance of ln(wages) for all workers due to wage variation between occupations will mechanically increase (from .201 in 2004 to .224 in 2019). In practice, because of changes in occupations over time,[8] I estimate the predicated variance of wages using average wages for each major occupational group that are estimated separately for each reference date. The portion of the variance of ln(wages) for all workers due to wage variation between occupations increases from 2004 (.204) through 2013 (.229), but then decreases from 2013 to 2019 (.203). The falling wage variation between occupations overall during the 2013 to 2019 period occurs because, as noted in Dey, Handwerker, Piccone, and Voorheis, there was particularly strong wage growth for lower-wage occupations in the OEWS (relative to other occupations) during these years.

---

[8] All occupations are recoded as consistently as possibly throughout this work, using crosswalk files created by the OEWS program for making tabulations from data collected before and after the 2010 and 2018 SOC revisions. However, additional changes in occupational definitions are not fully captured by these occupational recodings.

There is no mechanical relationship between overall changes in employment by occupation and the Herfindahl-Hirschman index: a version of the Herfindahl-Hirschman index that pools workers across all employers varies only between .0277 and .0283 over this period, with no clear time trend.

The actual time trend of mean occupational homogeneity at the establishment level is described with regressions of the form

(3) $$OccHomogeneity_{ijt} = \alpha Reference\ Date_t + \beta X_{ijt} + \varepsilon_{ijt}$$

where ReferenceDate measures time in decades since 2004, and X_{ijt} are other observable characteristics of individual *i* (occupation) and employer *j* (industry, geography, and size) at time *t*. Trend regression results for equation (3) are shown in Tables 2 and 3. The first two rows of Table 2 show an increase over time in the normalized Herfindahl-Hirschman measure of the occupational homogeneity of employers overall, but changes in occupations and employer characteristics explain about 75% of this increase. The partial predicted variance of ln(wages) measure of occupational homogeneity has fallen over time, also showing a trend of increasing employer homogeneity overall, with only 16% of this increase explained by changes in occupations and employer characteristics.

Further rows of Table 2 repeat this analysis for subgroups of occupations. Occupations (at the 3-digit minor occupational category SOC level) are grouped by average wage into quintiles, with roughly equal total weighted employment in each quintile. [9] Appendix A lists the occupations of each quintile, while counts of the observations for each quintile are in the Data

[9] To form quintiles, occupations are ranked by their average wages across all years. This grouping of occupations is quite stable over time.

Appendix. The list of occupations in the lowest-paid quintile is a short one, because the occupations in this quintile, such as Food and Beverage Serving Workers, tend to be large. The list of occupations in the highest-paid quintile, such as Social Scientists, is much longer, because these occupations tend to be smaller.

The subgroup rows of Table 2 show the greatest increases over time in the normalized Herfindahl-Hirschman measure of occupational homogeneity—after including controls for occupation and establishment characteristics—occur in the bottom and top quintiles of occupations. For the partial predicted variance measure of occupational homogeneity, the pattern is similar. The greatest decreases in this measure over time (indicating increases in occupational homogeneity)—after including controls for occupation and establishment characteristics—also occur in the lowest and highest paid quintiles of occupations.

Figure 2 uses the same five quintiles of occupations by typical wages used in Table 2, and shows the fraction of workers in each quintile of occupations who work in establishments without any workers in other quintiles. It is unsurprising that workers in all other quintiles of occupation are growing less likely to have any coworkers in the middle three quintiles, as the middle quintile occupations have declining shares of overall employment over time. However, Figure 2 shows that workers in the bottom three quintiles increasingly have no coworkers in the top quintile of occupation, and workers in upper four quintiles increasingly have no coworkers in the bottom quintile of occupation, although there has been no decline in employment in the bottom quintile of the occupational distribution and there has been an increasing share of employment in the top quintile of the occupational distribution over time. The polarization of employment is not happening evenly across establishments.

To illustrate the impact of these trends in employment by occupational quintiles on the predicted variance of wages for establishments, consider coarsening the occupational distribution into only three occupation groups: low-wage occupation group L, middle-wage occupation group M, and high-wage occupations H, with mean wages for occupations in each group $\overline{w_L} < \overline{w_M} < \overline{w_H}$ and within-occupations wage variances by group $\sigma_L^2 < \sigma_M^2 < \sigma_H^2$. Each establishment j contains $n_L \geq 0$ workers in the low-wage occupation group, $n_M \geq 0$ workers in the middle-wage occupation group, and $n_H \geq 0$ workers in the high-wage occupation group, with $n_L + n_M + n_H = n_J$. The predicted variance of wages for each establishment is $\widehat{V_J} = \frac{n_L\left[(\overline{w_L} - \widehat{\overline{w_J}})^2\right]}{n_j} + \frac{n_M\left[(\overline{w_M} - \widehat{\overline{w_J}})^2\right]}{n_j} + \frac{n_H\left[(\overline{w_H} - \widehat{\overline{w_J}})^2\right]}{n_j} + \frac{n_L {\sigma_L}^2}{n_j} + \frac{n_M {\sigma_M}^2}{n_j} + \frac{n_H {\sigma_H}^2}{n_j}$, and the first three terms are the "partial predicted variance" studied here.

For workers in occupation group L, employing establishments have higher n_L, lower n_M, and lower n_H, and, as shown in Figure 2, growing numbers of workers in occupation group L work in establishments with $n_M = n_H = 0$. There is little variation in wages between the low-wage occupations (a low value of ${\sigma_L}^2$), which reduces the typical values of $\left(\overline{w_L} - \widehat{\overline{w_J}}\right)$. With fewer workers in middle or high wage occupations, there is less weight on the other components of the predicted wage variance. This lowers $\widehat{V_J}$ for the workers in occupation group L.

For workers in occupation group H, employing establishments have lower n_L, lower n_M, and higher n_H, and, as shown in Figure 2, growing numbers of workers in occupation L work in establishments with $n_L = n_M = 0$. Although average wages are higher in these establishments, reducing the typical values of $(\overline{w_H} - \widehat{\overline{w_J}})$, the greater weight n_H associated with the high wage variance within these occupations, ${\sigma_H}^2$, means a greater $\widehat{V_J}$ overall for these establishments.

Weil (2014) describes how large corporations have shed many low-wage tasks by outsourcing them to other companies, which repeatedly subcontract them to smaller and smaller employers. Figure 3 shows that establishment size plays a role in the increasing segregation of workers in the lowest-paid quintile of occupations and workers in the highest-paid quintile of occupations into separate establishments, following the pattern Weil describes. Rising shares of employment for the lowest-paid quintile of occupations occurred only in establishments of less than 100 workers, while rising share of employment for the highest-paid quintile of occupations occurred more sharply in establishments of 100 or more workers.[10]

Table 3 shows the implication of this growing segregation of workers by establishment size for time trends in measured establishment occupational homogeneity. This table disaggregates the results of equation (3) by both establishment size and occupation group. For workers in the lowest-paid quintile of occupations in establishments with less than 100 workers, including the controls described above, workplaces are increasingly homogenous, by both measures. Furthermore, the predicted variance of ln(wages) measure of occupational

[10] Patterns are similar for establishments of 1-49 workers and establishments of 50-99 workers. Patterns are also quite similar when using EIN size instead of establishment size.

homogeneity has fallen faster (homogeneity has increased more) for low-paid workers in these smaller establishments than in establishments with 100 or more workers.

Appendices B-D describes employer occupational homogeneity trends when imputed data are not included, when defining employers by Employer Tax Identification Number (EIN) rather than establishments, and separately for states with high and low unionization rates.

Particularly for workers in low-wage occupations, all measures show a clear trend of increased employer occupational homogeneity over time. The next section shows these workers' wages are lower when they work for more occupationally homogenous employers.

IV: Relationships between Measured Occupational Homogeneity and Wages

The outsourcing literature provides several examples of occupations in which outsourcing is associated with lower wages, including occupations listed in Table 1. Among the example occupations in Table 1, all of the low wage occupations (food preparation and service, janitors, and security guards) earn considerably lower average wages in outsourced specialty industries than in other industries. These example occupations are examples precisely because there are obvious industries to which they can be outsourced; most other occupations do not have such obvious industries for outsourcing. However, the advantage of the occupational homogeneity measures in this paper is that they can be measured for the employers of all occupations. This section shows the relationship between occupational homogeneity and wages for all workers.

I describe the relationship between occupational homogeneity and wage with regressions of the form

$$\text{(4)} \qquad \ln(wage_{ijt}) = \alpha OccHomogeneity_{jt} + \beta X_{ijt} + \varepsilon_{ijt},$$

where $OccHomogeneity_{jt}$ is the measure of occupational homogeneity for the employer of individual *i* at employer *j* in time *t*, and X_{ijt} are other observable characteristics of individual *i* (occupation) and employer *j* (industry, geography, and size) at time *t*. Results of this regression are shown in Table 4. The first row of this table gives estimates of the impact of occupational homogeneity on wages, α, with no additional variables (other than a fixed effect for each reference date). These estimates show that increased occupational homogeneity is associated with lower wages overall. The second row of Table 4 gives these estimates with additional variables added to the regression. These detailed controls reduce the magnitude of the relationship between occupational homogeneity and wages, α, but the estimates maintain the same sign and remain very significant.

Further rows of Table 4 repeat this analysis for the same subgroups of occupations as in Table 2. The relationship between occupational homogeneity and wages, after controlling for own-occupation and employer characteristics, is generally stronger for workers in typically low- and middle-wage occupations than for workers in typically high-wage occupations. The relationship between the typical wage levels for a quintile of occupations and the wage coefficient of occupational homogeneity for the occupations in that quintile is not monotonic, with the largest wage coefficients for the quintile of occupations with the second-lowest typical wages.

There is one group of workers for whom greater occupational homogeneity—at least as measured by the predicted variance of wages between occupations—is associated with substantially *higher* wages, once own-occupation and employer characteristics are taken into account. These are the workers in the highest paid quintile of occupations. This is consistent with the model of Bilal and Lhuillier (2021), in which the outsourcing of lower-paid work is associated with greater demand—and higher wages—for work in higher-median-wage occupations.

Appendices B-D describe the relationship between occupational homogeneity and wages when imputed data are not included, when defining employers by Employer Tax Identification Number (EIN) rather than establishments, and separately for states with high and low unionization rates.

This section has described the relationship observed between occupational homogeneity and wages, but does not say employer homogeneity "causes" lower wages for workers in lower-wage occupations. The data used in this paper do not allow me to measure whether differences in unmeasured skills and tasks—within the same occupation—might explain some of the difference in wages between workers in more and less homogenous workplaces. For example, janitors who work in the janitorial services industry may lack some specialized skills of janitors in other industries, and may perform somewhat different tasks than those employed in other industries. However, the many U.S. examples described in Weil (2014) and the labor force histories of

German workers whose jobs are outsourced, as documented in Goldschmidt and Schmieder (2015) provide evidence that some portion of the observed relationship between employer homogeneity and wages is causal. The estimates in thus section should thus be considered an upper bound for the size of the causal impact of employer homogeneity on wages.

V. Occupational Homogeneity and Wage Inequality

The association between occupational homogeneity and lower wages—particularly for workers in lower-wage occupations—coupled with the trend of growing occupational homogeneity, suggests a role for occupational homogeneity in explaining growing wage inequality. Barth, Bryson, Davis, and Freeman (2016) highlighted that most inequality growth is between establishments, and is not explained by industry or geography. Moreover, Song, Price, Guvenen, and von Wachter (2019) show that the vast majority of pay-inequality growth at small and medium-sized firms in the United States from 1978-2013 was due to increasing segregation and sorting of workers who earn higher pay—without describing what about these workers makes them higher-paid workers—to firms that pay higher wages. Weil (2014) speculated that increased fissuring of employers could exacerbate wage inequality, but he did not have data to measure this directly. This section presents evidence showing that changes in occupational homogeneity contribute to the growth in wage inequality during this period.

I use Rios-Avila's implementation of Fortin, Firpo, and Lemiuex's Recentered Influence Function (RIF) Decomposition method to decompose changes in real ln(wage) variance from the

2004 reference date to the 2016 reference date[11] into portions that can be explained by the changing composition of workers by occupation, and the changing composition of their employing establishments by industry, geography, size, and occupational homogeneity. Because the occupational homogeneity measures are continuous rather than categorical variables, these variables are divided into quartiles for this reweighting exercise. The evidence in Table 2 shows that occupational homogeneity is changing in different ways for different quintiles of occupations. Thus, I interact occupational homogeneity variables with the same quintiles of occupation used above.[12] In addition, I add a dummy variable for lowest-wage quintile occupations employed establishment of less than 100 workers that are in the bottom half of the predicted variance distribution to the vector of indicator variables describing the predicted variance measure of occupational homogeneity.

Results are shown in Table 5. The changing composition of employment by industry, geography, establishment sizes, occupational quintiles, and the categories of occupational homogeneity described above can more than explain all of the change in ln(wage) variance from 2004 to 2016. Decomposing the change in ln(wage) variance by source, by far the category which most explains wage inequality growth is the changing pattern of employment by occupational quintiles (employment polarization), which explains 80% of wage inequality growth. Changes in employment by the occupational homogeneity of employers explains 13% of wage inequality growth—7% for changing values of the Hirschman-Herfindahl index, and 6%

[11] The 2016 reference date is chosen as the end date because the OEWS data show a sharp contraction in wage variance from 2016 to 2019, and so there is no overall wage variance growth to explain over the full 2004 to 2019 period. A more complete discussion of the fall in wage variance in the OEWS data in this period is in Dey, Piccone, Handwerker, and Voorheis (forthcoming).

[12] This follows the example of Goldschmidt and Schmieder (section V.C.), who use indicators for deciles of the firm wage effect interacted with dummies for frequently outsourced occupations.

for the changing values of predicted employer wage variance based on between-occupation wage variation. Changing industry composition explains much of the remaining growth in wage variation.

To examine the impact of changing occupational homogeneity on the growth of wage variance between establishments, I use the Dinardo-Fortin-Lemiuex (DFL) 1996 method. This method calculates counterfactual wage distributions by reweighting observable characteristics in the later period (2016) to their distributions in the earlier period (2004). The overall variance of real ln(wages) increased from 0.362 for the 2004 reference date to 0.380 for the 2016 reference date, and most of this increase is due to between-establishment wage variance increasing from 0.205 to 0.220. Reweighting the 2016 data to the 2004 distribution of employment by quartiles of both occupational homogeneity measures (without interacting these occupational homogeneity measures with occupational quintiles, so as to avoid also capturing the impact of employment polarization by occupation) and the indicator for workers in typically lower-wage occupations employed in small homogenous establishments, the between-establishment wage variance would be .211 rather than the actual .220. This reweighting explains about half of wage variance growth between establishments.

Wage variation, including the between-establishments portion of wage variation, declined from 2016 to 2019 (with the between-establishments portion of ln(wage) variance falling from .220 to .208). However, applying this reweighing method, the between establishments portion of ln (wage) variance in 2019 would have been still lower (.197) under the 2004 distribution of occupational homogeneity variables.[13]

[13] Reweighting 2016 or 2019 data to the 2004 distributions of occupational homogeneity variables without interacting these occupational homogeneity variables with occupations does not fully capture the impact of changes in occupational homogeneity, but has the advantage of not being co-mingled with changes in employment by occupation. This reweighting reduces the ln(wage) variance between establishments in the 2016 and 2019 data

In sum, these results show that changes in occupational homogeneity are a very important part of growing wage inequality for the lower 98.5% of the wage distribution. Both the Herfindahl-Hirschman measures of overall occupational homogeneity and the separation of typically-low wage occupations into separate workplaces from typically-high wage occupations are important for wage inequality growth during this period.

VI. Summary: Outsourcing and increasing wage inequality

While many authors have studied the growth in wage inequality between employers and others have studied the impact of outsourcing on wages in particular occupations and industries, this paper is among the first to connect the two with a study of the impact of the changing distribution of occupations between employers on wage inequality in the United States. This paper uses multiple measures of occupational homogeneity (at both the establishment and employer tax-ID levels) to examine the impact of outsourcing on wages and on wage inequality. These measures show greater occupational homogeneity for the occupations used to study outsourcing by Abraham and Taylor (1996); Dube and Kaplan (2010); Dey, Houseman, and Polivka (2010); and Goldschmidt and Schmeider (2015), when these occupations are employed in establishments in the outsourced sector. For example, employer occupational homogeneity is

without reducing overall ln(wage) variance. Reweighting 2016 or 2019 data to the 2004 distributions of occupational homogeneity variables and also to the 2004 distributions of the interactions of occupational homogeneity with occupational quintiles more completely captures the impact of changing occupational homogeneity in different parts of the wage distribution, but is co-mingled with changes in the occupational distribution from 2004 to 2016. This reduces ln(wage) variance between establishments even further (from .220 to .198 in 2016 and from .208 to .179 in 2019) and also reduces the overall ln(wage) variance (from .380 to .365 in 2016 and from .360 to .338 in 2019).

higher for janitors when they are employed in establishments in the janitorial services industry than when they are employed in other industries.

The advantage of measuring outsourcing with occupational homogeneity is that these measures can be calculated for every employee of every employer, not only for "case study" occupations. This paper shows that by two very different measures of occupational homogeneity—for employers of every size—there is an increase in employer homogeneity over time for the quintile of workers in the lowest-wage occupations. Falling employment levels for middle-wage occupations mean those in other occupations have fewer coworkers in middle-wage occupations, but low-wage workers also have a declining share over time of coworkers in high-wage occupations, even as high-wage occupations make up a growing share of employment. Low-wage occupations are growing in smaller employers, while the growth of high-wage occupations is concentrated in large employers. These patterns of time trends are consistent with the idea that in the economy as a whole, companies are "de-verticalizing" by outsourcing functions not part of their "core competencies," particularly if these outsourced tasks are done by workers paid lower wages than the "core workers" in the establishment.

The paper further shows that employer occupational homogeneity is related to wage levels. It has a particularly strong negative wage association for workers in occupations that are typically low paid, even after controlling for the occupations of employees and various observable characteristics of their employers. In contrast, workers in the highest paid quintile of occupations are paid more if they have fewer co-workers in typically low-wage occupations, after controlling for their own occupations and the observable characteristics of their employers.

Song, Price, Guvenen, and von Wachter (2019) show that the vast majority of pay-inequality growth at small and medium-sized firms is due to the increasing segregation and sorting of workers who earn lower pay—without describing what about these workers makes them lower-paid workers—to firms that pay lower wages. Occupation is just such a characteristic affecting workers' wages, and this paper shows that workers in low-wage occupations are increasingly concentrated at employers with fewer high-wage occupations, contributing to wage inequality growth.

Although the data used in this paper cannot show changes in the wage distribution for the very highest 1.3% of wage-earners, they are well suited to measure the contribution of employers' occupational homogeneity to wage inequality growth for the remaining 98.7% of the wage distribution. Decompositions of ln(wage) variance growth in these data show the growing polarization of employment can explain the vast majority of inequality growth, and the changing distribution of occupational homogeneity by the typical wage level of occupations can explain much of the remainder. Growing separation of workers in low-wage occupations from the employers of workers in high-wage occupations is an important component of recent wage inequality growth.

REFERENCES

Abowd, John M., Francis Kramarz, and David Margolis (1999). "High Wage Workers and High Wage Firms." *Econometrica*, pp. 251-334.

Abraham, Katharine G. and James R. Spletzer (2009). "New Evidence on the Returns to Job Skills." *American Economic Review Papers and Proceedings*, Vol. 99, No. 2, pp. 52-57.

Abraham, Katharine G. and James R. Spletzer (2010). "Are the New Jobs Good Jobs?," in *Labor in the New Economy*, edited by Katharine G. Abraham, James R. Spletzer, and Michael Harper. University of Chicago Press.

Abraham, Katherine G and Susan K. Taylor (1996). "Firms' use of Outside Contractors: Theory and Evidence." *Journal of Labor Economics,* 14(3): 394-424.

Acemoglu, Daron (2002). "Technical Change, Inequality, and the Labor Market." *Journal of Economic Literature*, Vol. 40, No. 1, pp. 7-72.

Autor, David, H., and David Dorn (2013). "The growth of low-skill service jobs and the polarization of the US labor market." *The American Economic Review*, Vol.103, No. 5, pp. 1553-1597.

Autor, David H., Lawrence F. Katz, and Melissa S. Kearney (2006). "The Polarization of the U.S. Labor Market," *American Economic Association Annual Meeting Papers and Proceedings*, pp. 189-194.

Autor, David H., Lawrence F. Katz, and Melissa S. Kearney (2008). "Trends in U.S. Wage Inequality: Re-Assessing the Revisionists." *Review of Economics and Statistics*, Vol. 90, No. 2, pp. 300-323.

Barth, Erling, Alex Bryson, James C. Davis, and Richard Freeman (2016). "It's Where You Work: Increases in Earnings Dispersion across Establishments and Individuals in the U.S." *Journal of Labor Economics,* 34(S2): S67-S97.

Bilal, Adrien, and Hugo Lhuillier. Outsourcing, Inequality and Aggregate Output. No. w29348. National Bureau of Economic Research, 2021.

Bound, John and George Johnson (1992). "Changes in the Structure of Wages in the 1980's: An Evaluation of Alternative Explanations." *The American Economic Review*, Vol. 82, No. 3, pp. 371-392.

Bronars, Stephen G. and Melissa Famulari (1997). "Wage, Tenure, and Wage Growth Variation Within and Across Establishments." *Journal of Labor Economics*, pp. 285-317.

Card, David, Jorg Heining, and Patrick Kline (2013). “Workplace Heterogeneity and the Rise of West German Wage Inequality.” *Quarterly Journal of Economics*, Vol. 128, No. 3, pp. 967-1015.

Dey, Matthew, David Piccone Jr., Elizabeth Weber Handwerker, and John Voorheis. “Were Wages Conversing During the 2010s Expansion?” *Monthly Labor Review*, forthcoming.

Dey, Matthew, and Elizabeth Weber Handwerker (2016). "Longitudinal data from the Occupational Employment and Wage Statistics survey." *Monthly Labor Review*.

Dey, Matthew, Susan Houseman, and Anne Polivka. 2010. "What Do We Know about Contracting Out in the United States? Evidence from Household and Establishment Surveys." In Labor in the New Economy, Katharine G. Abraham, James R. Spletzer, and Michael J. Harper, eds. Chicago: University of Chicago Press, 2010, pp. 267-304

Davis, Steve J. and John Haltiwanger (1991). “Wage Dispersion Between and Within U.S. Manufacturing Plants.” *Brookings Papers on Economic Activity*, pp. 115-200.

DiNardo, John, Nicole M. Fortin, and Thomas Lemieux (1996). “Labor Market Institutions and the Distribution of Wages, 1973-1992: A Semiparametric Approach.” *Econometrica*, Vol. 64, No. 5, pp. 1001-1044.

Drenik, Andres, Simon Jäger, Miguel Pascuel Plotkin, and Benjamin Schoefer (2021). Paying Outsourced Labor: Direct Evidence from Linked Temp Agency-Worker-Client Data. *The Review of Economics and Statistics*, pp. 1-28.

Dube, Arindrajit, and Ethan Kaplan (2010). "DOES outsourcing reduce wages in the low-wage service occupations? Evidence from janitors and guards." *ILR Review* 63.2: 287-306.

Dunne, Timothy, Lucia Foster, John Haltiwanger, and Kenneth R. Troske (2004). “Wage and Productivity Dispersion in United States Manufacturing: The Role of Computer Investment.” *Journal of Labor Economics*, Vol. 22, No. 2, pp. 397-429.

Employee Benefit Research Institute (EBRI) (2009). *Fundamentals of Employee Benefit Programs, 6th ed.* http://www.ebri.org/publications/books/index.cfm?fa=fundamentals

Erickcek, George, Susan Houseman, and Arne Kalleberg (2003). “The Effects of Temporary Services and Contracting Out on Low-Skilled Workers: Evidence from Auto Suppliers, Hospitals, and Public Schools,” in *Low-Wage America: How Employers are Reshaping Opportunity in the Workplace*, edited by Eileen Appelbaum, Annette Bernhardt, and Richard J. Murnane. New York: Russell Sage Foundation, pp.368-403.

Firpo, Sergio P., Nicole M. Fortin, and Thomas Lemieux (2018). "Decomposing Wage Distributions Using Recentered Influence Function Regressions." *Econometrics* 6 (3):41.

Fortin, Nicole, Thomas Lemieux, and Sergio Firpo (2011). "Decomposition Methods in Economics," *Handbook of labor economics* 4 (2011): 1-102.

Goldschmidt, Deborah and Johannes F. Schmieder (2017). "The Rise of Domestic Outsourcing and the Evolution of the German Wage Structure," *The Quarterly Journal of Economics* 132(3):1165-1217.

Goos, Maarten and Alan Manning (2007). "Lousy and Lovely Jobs: The Rising Polarization of Work in Britain," *Review of Economics and Statistics*, Vol. 89, No. 1, pp. 118-133.

Goos, Maarten, Alan Manning and Anna Salomons (2009). "Job Polarization in Europe," *American Economic Review Papers and Proceedings*, Vol. 99, No. 2, pp. 58-63.

Groshen, Erica L. (1991a). "Sources of Intra-Industry Wage Dispersion: How Much do Employers Matter?" *The Quarterly Journal of Economics*, Vol. 106, No. 3, pp. 869-884.

Groshen, Erica L. (1991b). "Five Reasons Why Wages Vary Among Employers." *Industrial Relations*, Vol. 30, No. 3, pp. 350-381.

Handwerker, Elizabeth Weber, & Mason, Lowell G. (2013). Linking firms with establishments in BLS microdata. *Monthly Labor Review*, Vol 136, No 14.

Handwerker, Elizabeth Weber and James R. Spletzer (2014). "Measuring the distribution of wages in the United States from 1996-2010 with the Occupational Employment Survey" *Monthly Labor Review*, May 2014

Handwerker, Elizabeth Weber, and James R. Spletzer (2016). "The Role of Establishments and the Concentration of Occupations in Wage Inequality." *Inequality: Causes and Consequences.* Emerald Group Publishing Limited, 2016. 167-193.

Juhn, Chinhui, Kevin M. Murphy, Brooks Pierce (1993). "Wage Inequality and the Rise in Returns to Skill Wage Inequality and the Rise in Returns to Skill." *The Journal of Political Economy*, Vol. 101, No. 3, pp. 410-442.

Katz, Lawrence F. and Autor, David H. (1999). "Changes in the Wage Structure and Earnings Inequality." In *Handbook of Labor Economics* (volume 3A), edited by Orley Ashenfelter and David Card, Elsevier, pp. 1463-1555.

Katz, Lawrence F. and Kevin M. Murphy (1992). "Changes in Relative Wages, 1963-1987: Supply and Demand Factors." *The Quarterly Journal of Economics*, Vol. 107, No. 1, pp. 35-78.

Lane, Julia. I., Salmon, Laurie A., & Spletzer, James R. (2007). Establishment wage differentials. *Monthly Labor Review*, Vol. 130,No.3.

Lemieux, Thomas (2006). “Increasing Residual Wage Inequality: Composition Effects, Noisy Data, or Rising Demand for Skill?” *American Economic Review*, Vol. 96, No. 3, pp. 461-498.

Lemieux, Thomas (2008). “The Changing Nature of Wage Inequality.” *Journal of Population Economics*, Vol. 21, No. 1, pp. 21-48.

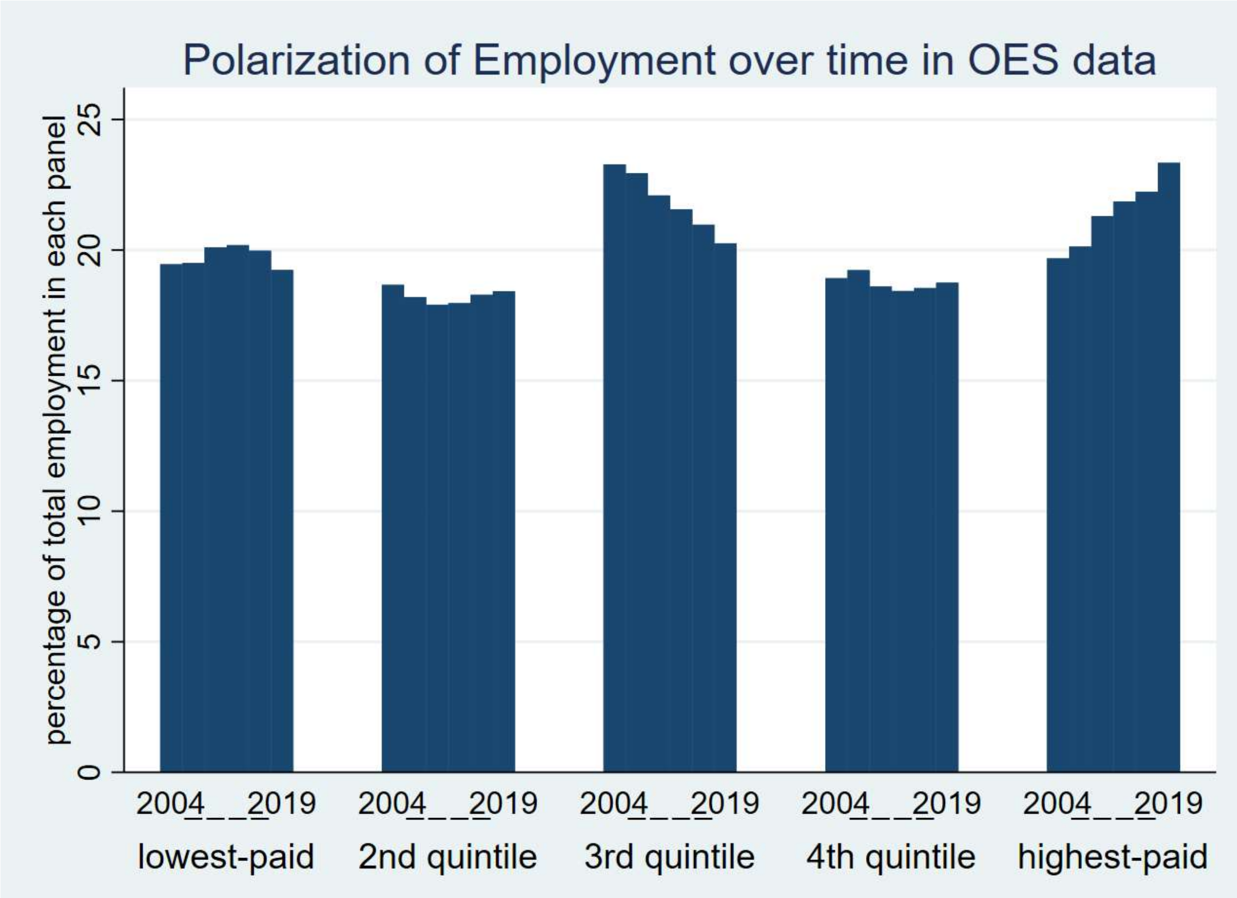
Perun, Pamela, "FAQs about Employees and Employee Benefits" (2010). *Georgetown Law School Memos and Fact Sheets*. Paper 46. Available at http://scholarship.law.georgetown.edu/legal/46

Rios-Avila, Fernando, “Recentered Influence Functions (RIF) in Stata: RIF Regression and RIF Decomposition” *Stata Journal.* Forthcoming.

Song, Jae, David J. Price, Fatih Guvenen, Nicholas Bloom, and Till von Wachter. (2019). "Firming up inequality." *The Quarterly Journal of Economics* Vol. 134, No 1, pp. 1-50.

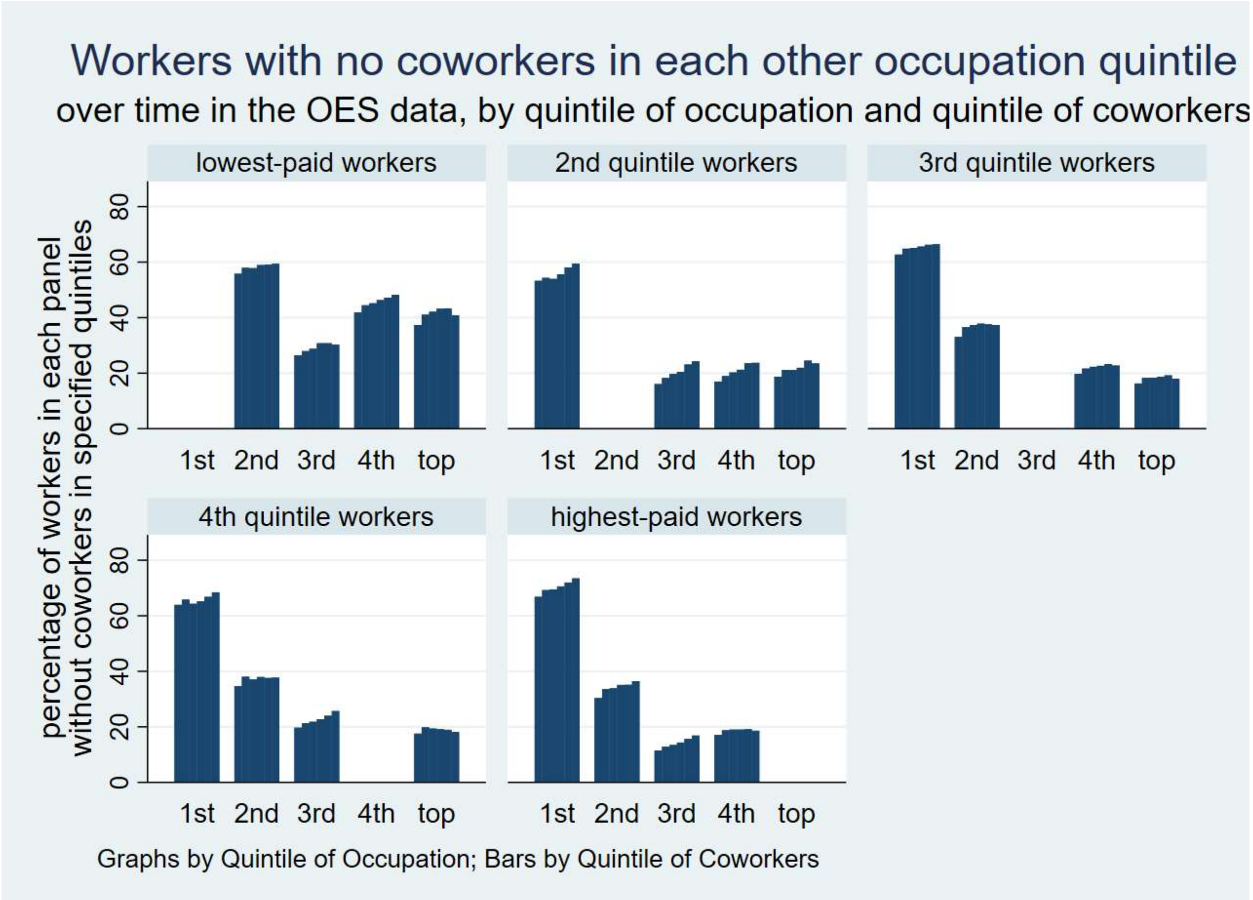
Weil, David. The fissured workplace. Harvard University Press, 2014.

Figure 1: Trends in Employment by Quintile of Occupation



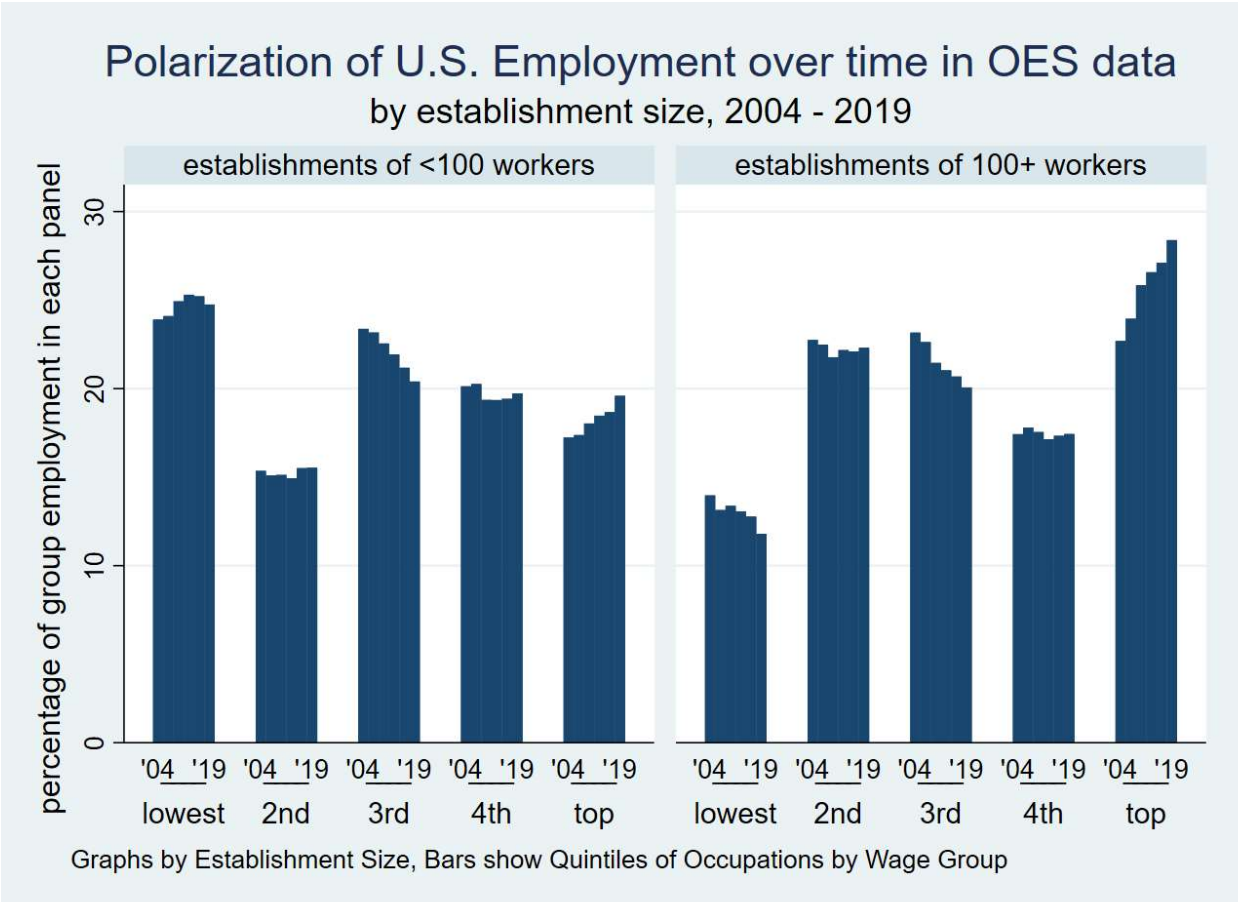
Note: The 94,928,505 employer-occupation-wage level observations from 6 panels of data are used to calculate overall average wage levels and employment levels. These are grouped into quintiles of minor occupational groups (3-digit SOC groups) by average wage levels (as shown in Appendix A). Quintiles may have slightly more or less than 20% of employment because of large occupational groups. This figure shows the percentage of employment in each occupational quintile in each panel of OEWS data, from November 2004 (collected from 2001 to 2004) through November 2019 (collected from 2017 to 2019).

Figure 2: Workers with no coworkers in other occupational quintiles over time in the OEWS data, by quintile of occupation and quintile of coworkers



Note: The 94,928,505 employer-occupation-wage level observations from 6 panels of data are used to calculate overall average wage levels and employment levels. These are grouped into quintiles of occupation by average occupational wage levels (as shown in Appendix A). This figure shows the percentage of workers in each quintile who are employed in establishments that have no workers in each other quintile, by panel (from November 2004 through November 2019). For example, the subgraph at the top left shows the fraction of workers in the lowest-quintile of occupations who have no co-workers in each other quintile of occupations, for each panel of the OEWS data.

Figure 3: Trends in Employment by Quintile of Occupation and Size of Employing Establishment



Note: The 94,928,505 employer-occupation-wage level observations from 6 panels of data are used to calculate overall average wage levels and employment levels. These are grouped into quintiles of minor occupational groups (3-digit SOC groups) by average wage levels (as shown in Appendix A). Quintiles may have slightly more or less than 20% of employment because of large occupational groups. This figure shows the percentage of employment in each establishment size group in each occupational quintile in each panel of OEWS data, from November 2004 (collected from 2001 to 2004) through November 2019 (collected from 2017 to 2019).

Table 1: Mean Values of Occupational Homogeneity for Specified Occupations and Industries, 2002-2016

Occupation and Industry	Avg ln(wage)	Mean Value of Occupational Homogeneity: Normalized Herfindahl of Occupational Homogeneity for the establishment	Mean Value of Occupational Homogeneity: Partial Predicted Variance of Wages for the establishment
Food preparation and serving (SOC 35)			
within Food Services (NAICS 722) – 81%	2.02	.464	.056
within all other industries – 19%	2.12	.237	.123
Janitors (SOC 372011)			
within Janitorial Services (NAICS 561720) –47%	2.09	.824	.042
within all other industries –53%	2.17	.286	.118
Security Guards (SOC 339032)			
within Security Guard Srvcs (NAICS 561612) –61%	2.20	.871	.029
within all other industries –39%	2.32	.314	.118
Truck Drivers (SOC 53303)			
within Truck Transportation (NAICS 484) –30%	2.68	.593	.039
within all other industries –70%	2.46	.339	.083
Accountants (SOC 132011)			
within Accounting Services (NAICS 541211) –25%	3.22	.485	.080
within all other industries –75%	3.16	.223	.132
Computer Occupations (SOC 151)			
within Computer Services (NAICS 5415) –28%	3.34	.500	.057
within all other industries –72%	3.29	.280	.115
Engineers (SOC 172)			
within Engineering Services (NAICS 54133) –21%	3.41	.320	.091
within all other industries –79%	3.43	.226	.124
Lawyers (SOC 231011)			
within Law Offices (NAICS 54111) –81%	3.76	.283	.277
within all other industries –19%	3.85	.227	.152

Notes: Data is pooled across 94,928,505 employer-occupation-wage level observations from 6 panels of data (with reference dates from November 2004 through November 2019). Each panel of data is collected over the previous three years. Normalized Herfindahl indices of Occupational Homogeneity for the establishment are calculated at the minor occupational group level, and are normalized for the number of employees in the establishment. Partial Predicted Variances of Wages for the establishment are based on employment by minor occupational group within the establishment, and do not include predicted within-occupational group wage variation.

Table 2: Change in mean values of Occupational Homogeneity over time

Occupational Homogeneity Variable	Normalized Herfindahl of Occupational Homogeneity for the establishment	Partial Predicted Variance of Wages for the establishment
All Occupations (94,628,505 observations)		
Raw Trend	0.00930	-0.00577
	(0.00002)	(0.00001)
All Controls	0.00246	-0.00485
	(0.00002)	(0.00000)
Lowest-paid quintile of occupations (7,725,178 observations)		
Raw Trend	0.01123	-0.00939
	(0.00004)	(0.00001)
All Controls	0.00488	-0.00712
	(0.00003)	(0.00001)
Second quintile of occupations (11,784,546 observations)		
Raw Trend	0.03088	-0.01020
	(0.00005)	(0.00001)
All Controls	0.00061	-0.00450
	(0.00004)	(0.00001)
Middle quintile of occupations (22,319,551 observations)		
Raw Trend	0.00904	-0.00277
	(0.00004)	(0.00001)
All Controls	-0.00035	-0.00109
	(0.00003)	(0.00001)
Fourth quintile of occupations (20,693,841 observations)		
Raw Trend	-0.00346	-0.00268
	(0.00004)	(0.00001)
All Controls	-0.00002	-0.00393
	(0.00004)	(0.00001)
Highest-paid quintile of occupations (32,105,389 observations)		
Raw Trend	0.00510	-0.00805
	(0.00004)	(0.00001)
All Controls	0.00443	-0.00703
	(0.00003)	(0.00001)

Note: These are coefficients α from regressions of the form $Occupation\ Homogeneity_{ijt} = \alpha Reference\ Date_t + \beta X_{ijt} + \varepsilon_{ijt}$, where the Reference Date is measured in decades since 2004 and X_{ijt} includes occupation fixed effects at the 6-digit SOC level, industry fixed effects at the 4-digit NAICS level, establishment size class, continuous establishment size, and state of location. Regressions are at the establishment-occupation-wage interval level, weighted by employment. Occupations (at the minor occupational group level) found in each quintile are listed in Appendix A. Normalized Herfindahl indices of Occupational Homogeneity for the establishment are calculated at the minor occupational group level, and are normalized for the number of employees in the establishment. Partial Predicted Variances of Wages for the establishment are based on employment by minor occupational group within the establishment, and do not include predicted within-occupational group wage variation. Standard errors are in parentheses.

Table 3: Change in mean values of Occupational Homogeneity over time, by establishment size

Occupational Homogeneity Variable	Normalized Herfindahl of Occupational Homogeneity for the establishment	Partial Predicted Variance of Wages for the establishment
Establishments with 1-99 Employees		
All Occupations (49,524,623 observations)		
All Controls	0.00394	-0.00585
	(0.00002)	(0.00001)
Lowest-paid quintile of occupations (4,922,494 observations)		
All Controls	0.00922	-0.00730
	(0.00004)	(0.00001)
Establishments with 100+ Employees		
All Occupations (45,103,882 observations)		
All Controls	-0.00079	-0.00327
	(0.00002)	(0.00001)
Lowest-paid quintile of occupations (2,802,684 observations)		
All Controls	-0.01233	-0.00570
	(0.00004)	(0.00001)

Note: These are coefficients α from regressions of the form $Occupation\ Homogeneity_{ijt} = \alpha Reference\ Date_t + \beta X_{ijt} + \varepsilon_{ijt}$, where the Reference Date is measured in decades since 2004 and X_{ijt} includes occupation fixed effects at the 6-digit SOC level, industry fixed effects at the 4-digit NAICS level, establishment size class, continuous establishment size, and state of location. All coefficients are significant at $p < 0.001$. Regressions are at the establishment-occupation-wage interval level, weighted by employment. Occupations (at the minor occupational group level) found in each quintile are listed in Appendix A. Normalized Herfindahl indices of Occupational Homogeneity for the establishment are calculated at the minor occupational group level, and are normalized for the number of employees in the establishment. Partial Predicted Variances of Wages for the establishment are based on employment by minor occupational group within the establishment, and do not include predicted within-occupational group wage variation. Standard errors in parentheses

Table 4: Regressions of log wages on measures of Occupational Homogeneity

Occupational Homogeneity Variable	Normalized Herfindahl of Occupational Homogeneity for the establishment	Partial Predicted Variance of Wages for the establishment
All Occupations (94,628,505 observations)		
With only date fixed effects	-0.530	1.830
	(0.000)	(0.000)
All Controls	-0.055	0.077
	(0.000)	(0.000)
Lowest-paid quintile of occupations (7,725,178 observations)		
With only date fixed effects	-0.171	0.631
	(0.000)	(0.000)
All Controls	-0.073	0.161
	(0.000)	(0.000)
Second quintile of occupations (11,784,546 observations)		
With only date fixed effects	-0.200	0.688
	(0.000)	(0.000)
All Controls	-0.077	0.341
	(0.000)	(0.000)
Middle quintile of occupations (22,319,551 observations)		
With only date fixed effects	-0.131	0.522
	(0.000)	(0.000)
All Controls	-0.070	0.300
	(0.000)	(0.000)
Fourth quintile of occupations (20,693,841 observations)		
With only date fixed effects	-0.188	0.525
	(0.000)	(0.001)
All Controls	-0.044	0.151
	(0.000)	(0.001)
Highest-paid quintile of occupations (32,105,389 observations)		
With only date fixed effects	-0.106	0.124
	(0.000)	(0.001)
All Controls	-0.014	-0.428
	(0.000)	(0.001)

Notes: These are coefficients α from regressions of the form $\ln(wage_{ijt}) = \alpha OccHomogeneity_{jt} + \beta X_{ijt} + \varepsilon_{ijt}$, where X includes reference date fixed effects, occupation fixed effects at the 6-digit SOC level, industry fixed effects at the 4-digit NAICS level, state fixed effects, and establishment size (using fixed effects for establishment size classes as well as continuous establishment size). All coefficients are significant at $p < 0.001$. Regressions are at the establishment-occupation-wage interval level, weighted by employment. Occupations (at the minor occupational group level) found in each quintile are listed in Appendix A. Normalized Herfindahl indices of Occupational Homogeneity for the establishment are calculated at the minor occupational group level, and are normalized for the number of employees in the establishment. Partial Predicted Variances of Wages for the establishment are based on employment by minor occupational group within the establishment, and do not include predicted within-occupational group wage variation. Standard errors in parentheses.

Table 5: Decomposition of Changes in real ln(wage) variance from 2004 to 2016

			Observations in late period (2016): 14,016,725 Observations in early period (2004): 13,012,513
Real log wage variance	Coeff.	Percent	Bootstrapped* Standard Deviations
Overall Variance			
Late period (2016 reference date)	0.3800		.0005
Counterfactual variance	0.3592		.0004
Early period (2004 reference date)	0.3621		.0006
Total change	0.0179	100%	.0007
of which explained (by compositional change)	0.0208	116%	.0001
of which unexplained (wage structure change)	-0.0029	-16%	.0007
Explained (compositional effect)			
Total	0.0208	100%	.0001
Pure explained	0.0208	100%	.0001
Specification error	0.0000	0%	.0000
Components of the pure explained effect			
Industry sector (2-digit NAICS)	0.0020	10%	.0001
Geography (Census Division)	0.0004	2%	.0000
Establishment size	-0.0009	-4%	.0000
Occupation quintiles (defined in Appendix A)	0.0167	80%	.0002
Normalized Herfindahl measure of establishments	0.0014	7%	.0001
Partial predicted variance of establishment ln(wages)	0.0012	6%	.0001
Unexplained (wage structure changes)			
Total	-0.0029	100%	.0007
Reweighting error	0.0000	-2%	.0003
Pure unexplained	-0.0029	102%	.0006

Notes: These are the results of Rios-Avila's implementation of Fortin, Firpo, and Lemiueux's Recentered Influence Function Decomposition. OEWS data with a 2004 reference date was collected from 2001 to 2004; date with a 2016 reference date was collected from 2014 to 2016. Industry here is grouped into 19 supersectors and geography into 7 Census divisions. Establishment size is measured in 9 categories (1-4 employees, 5-9, 10-19, 20-49, 50-99, 100-249, 250-499, 500-999, and 1000+). Quintiles of occupations are defined in Appendix A. Establishment-level normalized Herfindahl-Hirschman indices of occupations are measured with quartiles of the distribution, interacted with occupational quintiles. The predicted variance of ln(wage) for establishments is also divided into quartiles and interacted with occupational quintiles, with an additional dummy variable for low-wage occupations in establishment of less than 100 workers that are in the bottom half of the predicted variance distribution. *Standard deviations have not YET been bootstrapped—bootstrapping the coefficients with 300 replications is still in progress.

Data Appendix

This paper uses Occupational Employment and Wage Statistics (OEWS) Survey microdata. The OEWS survey is designed to measure occupational employment and wages in the United States by geography and industry, and is the only such survey of its size and scope, covering all establishments in the United States except those in agriculture, private households, and unincorporated self-employed workers without employees. Every year, approximately 400,000 private and local government establishments are asked to report the number of employees in each occupation paid within specific wage intervals: 200,000 establishments each November and another 200,000 each May. As described in Dey and Handwerker, the OEWS uses a complex sample design intended to minimize the variance of wage estimates for each occupation within industries and geographic areas. Thus, establishments expected to employ occupations with greater variation in wages have relatively larger probabilities of selection and lower estimation weights.

The OEWS survey form is a matrix of detailed occupations and wage intervals. For large establishments, the survey form lists 50 to 225 detailed occupations; these occupations pre-printed on the survey form are selected based on the industry and the size of the establishment. Small establishments write descriptions of the work done by their employees, which are coded into occupations by staff in state labor agencies. Wage intervals on the OEWS survey form are given in both hourly and annual nominal dollars, with annual earnings that are 2080 times the hourly wage rates. To calculate average wages, the OEWS program obtains the mean of each wage interval every year from the National Compensation Survey (NCS). These mean wages are then assigned to all employees in that wage interval. The OEWS survey is not designed to produce time series statistics.

The OEWS has been using the Standard Occupational Classification System since 1999, and had a change of industry classification systems from SIC to NAICS (2002) soon thereafter. Certain SOC and NAICS codes are combined to make groups consistent across the 2007 and 2012 NAICS revisions and the 2010 revision to the SOC. Data used in this paper begin in 2002 to avoid inconsistencies of SOC coding in small establishments during the initial years that the OEWS program used this coding system, as described by Abraham and Spletzer (2010).

Handwerker and Spletzer (2016) examine the decomposition of total wage variance in the OEWS into its within-establishment and between establishment components at length. Updating their findings, over the period of Fall 1999 through November 2016, 60% of wage variance is between establishments, while all of the growth in overall wage variance over this period is between establishments. Handwerker and Spletzer (2016) also find that similar amounts of establishment-level wage variance in the OEWS can be explained by broad industry groups to the amount found by Barth, Bryson, Davis, and Freeman. However, more of the establishment-level wage variance can be explained by detailed industry in the OEWS data than in the Census data, echoing findings comparing OEWS and CPS data.

Data Appendix Table 1: Summary Statistics

Variable	Observations	Employment represented	Weighted Mean	Min	Max	Standard Deviation
OEWS real wage	94,628,505	685,991,216	16.54	5.25	109.91	13.99
OEWS ln(wage)	94,628,505	685,991,216	2.59	1.66	4.70	0.61
Measured var(ln(wage)) of establishments	94,628,505	685,991,216	0.166	0.000	2.222	0.134
Herfindahl-Hirschman index of estab employment (by minor occupational group)	94,628,505	685,991,216	0.401	0.031	1.000	0.249
Normalized Herfindahl-Hirschman index of establishment employment	94,628,505	685,991,216	0.360	0.000	1.000	0.250
Predicted var of ln(wages) for the establishment (based on employment by minor occupational groups)	94,628,505	685,991,216	0.263	0.035	0.967	0.105
Portion of this predicted variance due to variation in wages **between** minor occupational groups	94,628,505	685,991,216	0.103	0.000	0.754	0.076
Establishment-level employment	94,628,505	685,991,216	559.63	1	56473	2128.64
Reference date for observation	94,628,505	685,991,216	2011.76	2004	2019	5.17
Decades since 2004	94,628,505	685,991,216	0.78	0.00	1.50	0.52

Variable Distributions	Observations	Employment represented	Fraction of Employment	Establishment observations
Quintiles of occupation – occupations are listed in Appendix A				
Lowest-paid quintile of occupations	7,725,178	135,364,590	19.7%	
2nd quintile of occupations	11,784,546	125,149,886	18.2%	
Middle quintile of occupations	22,319,551	149,537,779	21.8%	
4th quintile of occupations	20,693,841	128,587,364	18.7%	
Highest paid quintile of occupations	32,105,389	147,351,596	21.5%	
Major industry groups (2-digit)				
Agriculture, Forestry, Fishing and Hunting,	190,214	2,392,838	0.3%	30,319
Mining, Quarrying, and Oil and Gas Extraction	723,334	3,922,932	0.6%	42,031
Utilities	745,972	3,313,182	0.5%	36,539
Construction	5,165,398	40,525,748	5.9%	503,815
Manufacturing	15,633,177	76,728,433	11.2%	665,284
Wholesale Trade	6,550,147	34,747,748	5.1%	473,460
Retail Trade	9,858,784	92,921,634	13.5%	765,245
Transportation and Warehousing	2,898,187	28,394,498	4.1%	237,170
Information	3,531,304	17,192,071	2.5%	190,994
Finance and Insurance	5,954,353	38,345,572	5.6%	375,429
Real Estate and Rental and Leasing	1,376,340	9,043,767	1.3%	155,281
Professional, Scientific, and Technical Services	7,869,054	48,579,577	7.1%	587,715
Management of Companies and Enterprises	3,176,001	12,116,694	1.8%	73,799
Administrative and Support and Waste Management and Remediation Services	5,756,520	50,832,245	7.4%	440,223
Educational Services	2,853,391	15,726,201	2.3%	118,543
Health Care and Social Assistance	12,838,897	102,056,762	14.9%	688,914
Arts, Entertainment, and Recreation	2,116,635	12,294,224	1.8%	155,454

Variable Distributions	Observations	Employment represented	Fraction of Employment	Establishment observations
Accommodation and Food Services	4,050,790	73,089,407	10.7%	337,523
Major Occupational Groups (2-digit)				
Management Occupations	10,920,056	34,105,604	5.0%	
Business and Financial Operations Occupations	8,290,731	33,210,983	4.8%	
Computer and Mathematical Occupations	4,283,817	20,067,675	2.9%	
Architecture and Engineering Occupations	2,573,767	12,812,260	1.9%	
Life, Physical, and Social Science Occupations	907,683	4,294,966	0.6%	
Community and Social Service Occupations	1,167,878	7,156,326	1.0%	
Legal Occupations	520,072	4,659,370	0.7%	
Education, Training, and Library Occupations	1,407,548	12,134,228	1.8%	
Arts, Design, Entertainment, Sports, and Media Occupations	2,122,793	9,766,670	1.4%	
Healthcare Practitioners and Technical Occupations	4,132,160	38,777,463	5.7%	
Healthcare Support Occupations	1,542,717	28,748,772	4.2%	
Protective Service Occupations	651,463	7,377,864	1.1%	
Food Preparation and Serving Related Occupations	3,670,975	68,644,547	10.0%	
Building and Grounds Cleaning and Maintenance	2,118,482	21,712,311	3.2%	
Personal Care and Service Occupations	1,373,790	14,693,663	2.1%	
Sales and Related Occupations	8,488,583	84,341,624	12.3%	
Office and Administrative Support Occupations	19,330,700	101,129,136	14.7%	
Farming, Fishing, and Forestry Occupations	163,558	2,534,908	0.4%	
Construction and Extraction Occupations	3,091,111	32,378,384	4.7%	
Installation, Maintenance, and Repair Occupations	4,690,780	28,759,476	4.2%	
Production Occupations	6,987,796	54,476,434	7.9%	
Transportation and Material Moving Occupations	6,192,045	64,208,551	9.4%	

Appendix A: Occupations by Quintile

3-digit SOC code	*SOC Title*	*Average ln(wage)*	*Cummulative percentage of employment*	*Occupation Quintile*
359	Other Food Preparation and Serving Related Workers	1.97	1.1%	1
353	Food and Beverage Serving Workers	1.98	6.7%	1
393	Entertainment Attendants and Related Workers	2.01	7.2%	1
352	Cooks and Food Preparation Workers	2.08	9.6%	1
452	Agricultural Workers	2.08	9.9%	1
412	Retail Sales Workers	2.11	17.4%	1
392	Animal Care and Service Workers	2.11	17.5%	1
372	Building Cleaning and Pest Control Workers	2.12	19.7%	1
311	Nursing, Psychiatric, and Home Health Aides	2.14	22.7%	2
516	Textile Apparel and Furnishings Workers	2.16	23.2%	2
536	Other Transportation Workers	2.17	23.5%	2
399	Other Personal Care and Service Workers	2.19	24.3%	2
396	Baggage Porters Bellhops and Concierges	2.22	24.4%	2
395	Personal Appearance Workers	2.23	24.8%	2
537	Material Moving Workers	2.23	30.3%	2
397	Tour and Travel Guides	2.23	30.3%	2
373	Grounds Maintenance Workers	2.24	31.0%	2
339	Other Protective Service Workers	2.25	32.0%	2
513	Food Processing Workers	2.25	32.6%	2
432	Communications Equipment Operators	2.29	32.8%	2
259	Other Education, Training, and Library Occupations	2.31	33.1%	2
473	Helpers Construction Trades	2.32	33.3%	2
453	Fishing and Hunting Workers	2.34	33.3%	2
517	Woodworkers	2.36	33.6%	2
439	Other Office and Administrative Support Workers	2.36	36.4%	2
512	Assemblers and Fabricators	2.39	38.0%	2
434	Information and Record Clerks	2.40	42.2%	3
519	Other Production Occupations	2.42	44.6%	3
319	Other Healthcare Support Occupations	2.43	45.8%	3
351	Supervisors of Food Preparation and Serving Workers	2.46	46.6%	3
433	Financial Clerks	2.49	49.2%	3
533	Motor Vehicle Operators	2.49	52.2%	3
435	Material, Recording, Scheduling, Dispatching, and Distributing Workers	2.49	53.5%	3
332	Fire Fighting and Prevention Workers	2.49	53.5%	3
515	Printing Workers	2.51	53.7%	3
454	Forest Conservation and Logging Workers	2.53	53.8%	3
253	Other Teachers and Instructors	2.55	54.1%	3
252	Preschool, Primary, Secondary, and Special Education School Teachers	2.55	54.7%	3
514	Metal Workers and Plastic Workers	2.56	56.3%	3

3-digit SOC code	*SOC Title*	*Average ln(wage)*	*Cummulative percentage of employment*	*Occupation Quintile*
436	Secretaries and Administrative Assistants	2.57	58.9%	3
394	Funeral Service Workers	2.57	58.9%	3
419	Other Sales and Related Workers	2.58	59.6%	3
391	Supervisors of Personal Care and Service Workers	2.58	59.8%	3
333	Law Enforcement Workers	2.59	59.8%	3
211	Counselors, Social Workers, and Other Community and Social Service Specialists	2.60	60.8%	4
371	Supervisors of Building and Grounds Cleaning and Maintenance Workers	2.63	61.0%	4
493	Vehicle and Mobile Equipment Mechanics, Installers, and Repairers	2.63	62.2%	4
499	Other Installation Maintenance and Repair Occupations	2.65	64.3%	4
292	Health Technologists and Technicians	2.67	66.3%	4
212	Religious Workers	2.67	66.4%	4
312	Occupational Therapy and Physical Therapist Assistants and Aides	2.68	66.5%	4
475	Extraction Workers	2.69	66.7%	4
472	Construction Trades Workers	2.71	70.4%	4
274	Media and Communication Equipment Workers	2.71	70.5%	4
474	Other Construction and Related Workers	2.73	70.7%	4
451	Supervisors of Farming, Fishing, and Forestry Workers	2.73	70.7%	4
411	Supervisors of Sales Workers	2.75	72.0%	4
271	Art and Design Workers	2.76	72.4%	4
272	Entertainers and Performers, Sports and Related Workers	2.76	72.8%	4
331	Supervisors of Protective Service Workers	2.77	72.8%	4
194	Life, Physical, and Social Science Technicians	2.77	73.0%	4
299	Other Healthcare Practitioners and Technical Occupations	2.78	73.1%	4
254	Librarians, Curators, and Archivists	2.79	73.1%	4
492	Electrical and Electronic Equipment Mechanics, Installers, and Repairers	2.81	73.6%	4
232	Legal Support Workers	2.83	73.9%	4
531	Supervisors of Transportation and Material Moving Workers	2.87	74.2%	4
431	Supervisors of Office and Administrative Support Workers	2.89	75.3%	4
535	Water Transportation Workers	2.91	75.3%	4
173	Drafters, Engineering Technicians, and Mapping Technicians	2.91	75.9%	4
273	Media and Communication Workers	2.93	76.3%	4
511	Supervisors of Production Workers	2.98	76.8%	4

3-digit SOC code	SOC Title	Average ln(wage)	Cummulative percentage of employment	Occupation Quintile
534	Rail Transportation Workers	2.98	76.9%	4
413	Sales Representatives: Services	2.99	78.4%	4
518	Plant and System Operators	3.02	78.5%	4
414	Sales Representatives: Wholesale and Manufacturing	3.06	80.1%	5
491	Supervisors of Installation, Maintenance, and Repair Workers	3.08	80.4%	5
471	Supervisors of Construction and Extraction Workers	3.10	80.8%	5
131	Business Operations Specialists	3.10	83.8%	5
195	Occupational Health and Safety Specialists and Technicians	3.11	83.8%	5
193	Social Scientists and Related Workers	3.18	83.9%	5
132	Financial Specialists	3.18	85.8%	5
171	Architects, Surveyors, and Cartographers	3.18	86.0%	5
251	Postsecondary Teachers	3.20	86.4%	5
532	Air Transportation Workers	3.23	86.6%	5
151	Computer Specialists	3.30	89.4%	5
192	Physical Scientists	3.32	89.6%	5
191	Life Scientists	3.32	89.7%	5
119	Other Management Occupations	3.34	91.1%	5
152	Mathematical Science Occupations	3.35	91.2%	5
291	Health Diagnosing and Treating Practitioners	3.38	94.8%	5
172	Engineers	3.43	96.0%	5
111	Top Executives	3.61	97.8%	5
113	Operations Specialties Managers	3.63	99.0%	5
112	Advertising, Marketing, Promotions, Public Relations, and Sales Managers	3.67	99.6%	5
231	Lawyers, Judges, and Related Workers	3.75	100.0%	5

Appendix B: Dropping Imputations

The general practice in the wage inequality literature based on the Current Population Survey, such as Lemieux (2006), is to drop imputed data in the analysis. However, the imputations in the Occupational Employment and Wage Statistics microdata are an integral part of the estimation strategy for official publications based on this survey. They are constructed with a great deal of information, using nearest-neighbor matching with separate procedures for employment and wage variables. The estimation weights assume the inclusion of the imputed data; the imputation procedures are essentially more detailed weights on non-imputed data. However, in this Appendix, I check that the main results in this paper are robust to dropping imputed data.

As shown below, the results are largely consistent with those in tables 2, 4, and 5. Table B1, like Table 2, shows overall increases in occupational homogeneity by the Herfindahl-Hirschman measure, as well as increases by this measure in the lowest and highest paid quintile of occupations with and without controls, and for every quintile with controls except the 4th quintile. This Table also shows increases in occupational homogeneity by the partial predicted variance of wages measure overall and for all quintiles of the occupational distribution, with and without additional controls. Table B2, like Table 4, shows that by each measure of occupational homogeneity, overall and for all quintiles of the occupational distribution, with and without additional controls, greater occupational homogeneity is associated with higher wages—except for the top quintile of occupations by the predicted variance of wages between occupations measure, with the additional controls. Table B3, like Table 5, shows that changes in the occupational homogeneity of establishments from 2004 to 2016 contribute substantially to increases in overall wage variance between these two reference dates.

Appendix Table B1: Change in mean values of Occupational Homogeneity over time, dropping imputed data

Occupational Homogeneity Variable	Normalized Herfindahl of Occupational Homogeneity for the establishment	Partial Predicted Variance of Wages for the establishment
All Occupations (57,639,130 observations)		
Raw Trend	0.00453	-0.00367
	(0.00002)	(0.00001)
All Controls	0.00172	-0.00426
	(0.00002)	(0.00001)
Lowest-paid quintile of occupations (5,031,647 observations)		
Raw Trend	0.00577	-0.00682
	(0.00005)	(0.00001)
All Controls	0.00465	-0.00618
	(0.00004)	(0.00001)
Second quintile of occupations (7,304,452 observations)		
Raw Trend	0.02605	-0.00940
	(0.00006)	(0.00002)
All Controls	-0.00122	-0.00401
	(0.00005)	(0.00001)
Middle quintile of occupations (13,674,543 observations)		
Raw Trend	0.00605	-0.00110
	(0.00005)	(0.00001)
All Controls	-0.00138	-0.00040
	(0.00004)	(0.00001)
Fourth quintile of occupations (12,869,598 observations)		
Raw Trend	-0.00651	-0.00108
	(0.00005)	(0.00002)
All Controls	-0.00012	-0.00342
	(0.00005)	(0.00001)
Highest-paid quintile of occupations (18,758,890 observations)		
Raw Trend	0.00219	-0.00604
	(0.00004)	(0.00002)
All Controls	0.00423	-0.00671
	(0.00004)	(0.00001)

Note: These are coefficients α from regressions of the form $Occupation\ Homogeneity_{ijt} = \alpha Reference\ Date_t + \beta X_{ijt} + \varepsilon_{ijt}$, where the Reference Date is measured in decades since 2004 and X_{ijt} includes occupation fixed effects at the 6-digit SOC level, industry fixed effects at the 4-digit NAICS level, establishment size class, continuous establishment size, and state of location. Regressions are at the establishment-occupation-wage interval level, weighted by employment. Occupations (at the minor occupational group level) found in each quintile are listed in Appendix A. Normalized Herfindahl indices of Occupational Homogeneity for the establishment are calculated at the minor occupational group level, and are normalized for the number of employees in the establishment. Partial Predicted Variances of Wages for the establishment are based on employment by minor occupational group within the establishment, and do not include predicted within-occupational group wage variation. Standard errors are in parentheses.

Appendix Table B2: Regressions of log wages on measures of Occupational Homogeneity, dropping imputed data

Occupational Homogeneity Variable	Normalized Herfindahl of Occupational Homogeneity for the establishment	Partial Predicted Variance of Wages for the establishment
All Occupations (57,639,130 observations)		
With only date fixed effects	-0.539	1.880
	(0.000)	(0.000)
All Controls	-0.074	0.124
	(0.000)	(0.000)
Lowest-paid quintile of occupations (5,031,647 observations)		
With only date fixed effects	-0.208	0.717
	(0.000)	(0.000)
All Controls	-0.097	0.236
	(0.000)	(0.000)
Second quintile of occupations (7,304,452 observations)		
With only date fixed effects	-0.215	0.788
	(0.000)	(0.000)
All Controls	-0.102	0.447
	(0.000)	(0.001)
Middle quintile of occupations (13,674,543 observations)		
With only date fixed effects	-0.149	0.596
	(0.000)	(0.001)
All Controls	-0.094	0.408
	(0.000)	(0.001)
Fourth quintile of occupations (12,869,598 observations)		
With only date fixed effects	-0.197	0.587
	(0.000)	(0.001)
All Controls	-0.060	0.215
	(0.000)	(0.001)
Highest-paid quintile of occupations (18,758,890 observations)		
With only date fixed effects	-0.122	0.145
	(0.000)	(0.001)
All Controls	-0.021	-0.541
	(0.000)	(0.001)

Notes: These are coefficients α from regressions of the form $\ln(wage_{ijt}) = \alpha OccHomogeneity_{jt} + \beta X_{ijt} + \varepsilon_{ijt}$, where X includes reference date fixed effects, occupation fixed effects at the 6-digit SOC level, industry fixed effects at the 4-digit NAICS level, state fixed effects, and establishment size (using fixed effects for establishment size classes as well as continuous establishment size). All coefficients are significant at $p < 0.001$. Regressions are at the establishment-occupation-wage interval level, weighted by employment. Normalized Herfindahl indices of Occupational Homogeneity for the establishment are calculated at the minor occupational group level, and are normalized for the number of employees in the establishment. Partial Predicted Variances of Wages for the establishment are based on employment by minor occupational group within the establishment, and do not include predicted within-occupational group wage variation. Standard errors in parentheses.

Appendix Table B3: Decomposition of Changes in real ln(wage) variance from 2004 to 2016, dropping imputed data

			Observations in late period (2016): 8,366,419
			Observations in early period (2004): 8,476,264
Real log wage variance	Coeff.	Percent	Standard Deviations
Overall Variance			
Late period (2016 reference date)	0.3763		0.0006
Counterfactual variance	0.3538		0.0005
Early period (2004 reference date)	0.3588		0.0007
Total change	0.0174	100%	0.0009
of which explained (by compositional change)	0.0224	129%	0.0001
of which unexplained (wage structure change)	-0.0050	-29%	0.0009
Explained (compositional effect)			
Total	0.0224	100%	0.0001
Pure explained	0.0223	100%	0.0002
Specification error	0.0001	0%	0.0000
Components of the pure explained effect			
Industry sector (2-digit NAICS)	0.0025	11%	0.0001
Geography (Census Division)	0.0007	3%	0.0000
Establishment size	-0.0005	-2%	0.0000
Occupation quintiles (defined in Appendix A)	0.0168	75%	0.0003
Normalized Herfindahl measure of establishments	0.0020	9%	0.0002
Partial predicted variance of establishment ln(wages)	0.0009	4%	0.0001
Unexplained (wage structure changes)			
Total	-0.0050	100%	0.0009
Reweighting error	0.0000	1%	0.0004
Pure unexplained	-0.0050	99%	0.0008

Notes: These are the results of Rios-Avila's implementation of Fortin, Firpo, and Lemiueux's Recentered Influence Function Decomposition. OEWS data with a 2004 reference date was collected from 2001 to 2004; date with a 2016 reference date was collected from 2014 to 2016. Industry here is grouped into 19 supersectors and geography into 7 Census divisions. Establishment size is measured in 9 categories (1-4 employees, 5-9, 10-19, 20-49, 50-99, 100-249, 250-499, 500-999, and 1000+). Quintiles of occupations are defined in Appendix A. Establishment-level normalized Herfindahl-Hirschman indices of occupations are measured with quartiles of the distribution, interacted with occupational quintiles. The predicted variance of ln(wage) for establishments is also divided into quartiles and interacted with occupational quintiles, with an additional dummy variable for low-wage occupations in establishment of less than 100 workers that are in the bottom half of the predicted variance distribution. Standard deviations have not been bootstrapped.

Appendix C: Measuring employer size and homogeneity using EINs instead of establishments

Song, Price, Guvenen, Bloom, and von Wachter (2019) argue that the unit of importance for wage inequality should be the firm and not the establishment. In thinking about occupational homogeneity, some of the reasons for employers to outsource work to other establishments are also reasons to outsource work to other employers entirely. It may be more efficient for even multi-establishment employers to specialize in particular areas of work. Regulatory incentives for multi-establishment employers to specialize in employing workers in a particular part of the wage distribution are less clear. ERISA laws define employers as “controlled groups of corporations” and “entities under common control” in requiring common levels of pension and welfare benefits among most employees in exchange for favorable tax treatment (Perun 2010), and the Affordable Care Act of 2010 extended these provisions by requiring common levels of health care benefits among most employees of businesses with a common owner. However, as Perun notes, “Employers often invent new organizational structures and worker classifications designed to limit participation to favored employees… Regulatory authorities in turn develop complicated rules and regulations designed to prevent this.”

This paper focuses on measures of occupational homogeneity at the establishment level because establishments are the sampling units of the OEWS, and the OEWS sampling design often includes some but not all establishments of multi-establishment companies, particularly when there are establishments in geographic areas with fewer establishments available to sample. However, the OEWS microdata can be linked with the EIN (tax-ID) numbers that these establishments submit to the unemployment insurance system, as compiled by the Quarterly Census of Employment and Wages. As discussed extensively in Handwerker and Mason (2013), very large firms may use multiple EINs in the unemployment insurance system, and linking together all of the establishments in these data for very large firms involves a tremendous amount of manual review. Thus, while it is straightforward to recalculate measures of occupational homogeneity at the EIN level and repeat the analyses above, the reader should be cautioned that such EIN-level measures are not true firm-level measures.

OEWS data show that workers in the bottom quintile of occupations were paid more in huge firms than in smaller firms during earlier waves of data collection, but this difference disappeared around November 2013. This is consistent with the finding of Song et. al. that workers with low values of worker fixed effects in very large firms have seen declining wages over time. It is not exactly comparable to Song et. al. because those authors use repeated observations of workers over time to estimate worker fixed effects, an estimation not possible with the OEWS data. However, there is likely a great deal of overlap between workers in typically-low-wage occupations and workers with low fixed effects.

As shown below, the results are largely consistent with those in tables 2 and 4. Table B1, like Table 2, shows overall increases in occupational homogeneity by the Herfindahl-Hirschman measure, as well as increases by this measure in the lowest and highest paid quintile of occupations with and without controls, and for every quintile with controls except the 4th quintile. This Table also shows increases in occupational homogeneity by the partial predicted variance of wages measure overall and for all quintiles of the occupational distribution, with and without additional controls. Table B2, like Table 4, shows that by each measure of occupational homogeneity, overall and for all quintiles of the occupational distribution, with and without additional controls, greater occupational homogeneity is associated with higher wages—except for the top quintile of occupations by the predicted variance of wages between occupations measure, with the additional controls. However, Table B3, does not show the same role for increased occupational homogeneity in increasing overall wage variance between the 2004 and 2016 reference dates. This is because the predicted wage variance between occupations employed by employers at the EIN level in 2016 was slightly higher than it had been in 2004, before dropping sharply from 2016 to 2019 (as reflected in the overall trends shown in table B1).

Appendix Table C1: Change in mean values of EIN-level Occupational Homogeneity over time

Occupational Homogeneity Variable	Normalized Herfindahl of Occupational Homogeneity for the EIN	Partial Predicted Variance of Wages for the EIN
All Occupations (94,628,505 observations)		
Raw Trend	0.00625	-0.00418
	(0.00002)	(0.00001)
All Controls	0.00039	-0.00363
	(0.00001)	(0.00000)
Lowest-paid quintile of occupations (7,725,178 observations)		
Raw Trend	0.00467	-0.00913
	(0.00004)	(0.00001)
All Controls	-0.00189	-0.00661
	(0.00003)	(0.00001)
Second quintile of occupations (11,784,546 observations)		
Raw Trend	0.03295	-0.00840
	(0.00005)	(0.00001)
All Controls	0.00201	-0.00334
	(0.00003)	(0.00001)
Middle quintile of occupations (22,319,551 observations)		
Raw Trend	0.00667	-0.00177
	(0.00003)	(0.00001)
All Controls	-0.00069	-0.00028
	(0.00003)	(0.00001)
Fourth quintile of occupations (20,693,841 observations)		
Raw Trend	-0.00828	-0.00066
	(0.00004)	(0.00001)
All Controls	-0.00192	-0.00284
	(0.00003)	(0.00001)
Highest-paid quintile of occupations (32,105,389 observations)		
Raw Trend	0.00096	-0.00517
	(0.00003)	(0.00001)
All Controls	0.00201	-0.00481
	(0.00003)	(0.00001)

Note: These are coefficients α from regressions of the form $Occupation\ Homogeneity_{ijt} = \alpha Reference\ Date_t + \beta X_{ijt} + \varepsilon_{ijt}$, where the Reference Date is measured in decades since 2004 and X_{ijt} includes occupation fixed effects at the 6-digit SOC level, industry fixed effects at the 4-digit NAICS level, EIN size class, continuous EIN size, and state of location. Regressions are at the establishment-occupation-wage interval level, weighted by employment. Occupations (at the minor occupational group level) found in each quintile are listed in Appendix A. Normalized Herfindahl indices of Occupational Homogeneity for the establishment are calculated at the minor occupational group level, and are normalized for the number of employees in the EIN. Partial Predicted Variances of Wages for the establishment are based on employment by minor occupational group within the EIN, and do not include predicted within-occupational group wage variation. Standard errors are in parentheses.

Appendix Table C2: Regressions of log wages on EIN-level measures of Occupational Homogeneity

Occupational Homogeneity Variable	Normalized Herfindahl of Occupational Homogeneity for the EIN	Partial Predicted Variance of Wages for the EIN
All Occupations (94,628,505 observations)		
With only date fixed effects	-0.569	2.064
	(0.000)	(0.000)
All Controls	-0.052	0.127
	(0.000)	(0.000)
Lowest-paid quintile of occupations (7,725,178 observations)		
With only date fixed effects	-0.174	0.671
	(0.000)	(0.000)
All Controls	-0.071	0.183
	(0.000)	(0.000)
Second quintile of occupations (11,784,546 observations)		
With only date fixed effects	-0.230	0.755
	(0.000)	(0.000)
All Controls	-0.084	0.358
	(0.000)	(0.000)
Middle quintile of occupations (22,319,551 observations)		
With only date fixed effects	-0.136	0.537
	(0.000)	(0.000)
All Controls	-0.073	0.337
	(0.000)	(0.001)
Fourth quintile of occupations (20,693,841 observations)		
With only date fixed effects	-0.223	0.684
	(0.000)	(0.001)
All Controls	-0.033	0.174
	(0.000)	(0.001)
Highest-paid quintile of occupations (32,105,389 observations)		
With only date fixed effects	-0.130	0.350
	(0.000)	(0.001)
All Controls	-0.003	-0.325
	(0.000)	(0.001)

Notes: These are coefficients α from regressions of the form $\ln(wage_{ijt}) = \alpha OccHomogeneity_{jt} + \beta X_{ijt} + \varepsilon_{ijt}$, where Date X includes reference date fixed effects, occupation fixed effects at the 6-digit SOC level, industry fixed effects at the 4-digit NAICS level, state fixed effects, and EIN size (using fixed effects for EIN size classes as well as continuous EIN size). All coefficients are significant at $p < 0.001$. Regressions are at the establishment-occupation-wage interval level, weighted by employment. Normalized Herfindahl indices of Occupational Homogeneity for the EIN are calculated at the minor occupational group level, and are normalized for the number of employees in the EIN. Partial Predicted Variances of Wages for the EIN are based on employment by minor occupational group within the EIN, and do not include predicted within-occupational group wage variation. Standard errors in parentheses.

Appendix Table C3: Decomposition of Changes in real ln(wage) variance from 2002-2003 to 2014-2015, using EIN-level measures of Occupational Homogeneity and employer size

	Observations in late period (2016): 14,016,725		
	Observations in early period (2004): 13,012,513		
Real log wage variance	Coeff.	Percent	Standard Deviations
Overall Variance			
Late period (2016 reference date)	0.3800		0.0005
Counterfactual variance	0.3589		0.0004
Early period (2004 reference date)	0.3621		0.0006
Total change	0.0179	100%	0.0007
of which explained (by compositional change)	0.0211	118%	0.0001
of which unexplained (wage structure change)	-0.0032	-18%	0.0007
Explained (compositional effect)			
Total	0.0211	100%	0.0001
Pure explained	0.0209	99%	0.0001
Specification error	0.0002	1%	0.0000
Components of the pure explained effect			
Industry sector (2-digit NAICS)	0.0018	9%	0.0001
Geography (Census Division)	0.0004	2%	0.0000
EIN size	-0.0002	-1%	0.0000
Occupation quintiles (defined in Appendix A)	0.0179	86%	0.0002
Normalized Herfindahl measure of establishments	0.0013	6%	0.0001
Partial predicted variance of establishment ln(wages)	-0.0004	-2%	0.0001
Unexplained (wage structure changes)			
Total	-0.0032	100%	0.0007
Reweighting error	0.0000	1%	0.0003
Pure unexplained	-0.0032	99%	0.0006

Notes: These are the results of Rios-Avila's implementation of Fortin, Firpo, and Lemiueux's Recentered Influence Function Decomposition. OEWS data with a 2004 reference date was collected from 2001 to 2004; date with a 2016 reference date was collected from 2014 to 2016. Industry here is grouped into 19 supersectors and geography into 7 Census divisions. EIN size is measured in 10 categories (1-4 employees, 5-9, 10-19, 20-49, 50-99, 100-249, 250-499, 500-999, 1,000-9,999, and 10,000+). Quintiles of occupations are defined in Appendix A. Establishment-level normalized Herfindahl-Hirschman indices of occupations are measured with quartiles of the distribution, interacted with occupational quintiles. The predicted variance of ln(wage) for establishments is also divided into quartiles and interacted with occupational quintiles, with an additional dummy variable for low-wage occupations in establishment of less than 100 workers that are in the bottom half of the predicted variance distribution. Standard deviations have not been bootstrapped.

Appendix D: Heterogeneity by state-level unionization rates

One factor which may impact both wages and the organization of production (including the variety of occupations at a workplace) is unionization. The OEWS does not collect information on unionization patterns by employer, but it includes location of each establishment, and unionization rates vary strongly by state. Thus, state-level unionization rates are used to group the data into highly unionized states (17-26% of employed workers unionized), middle, and low unionized states (3-9.3% unionized), based on published tables from the Current Population Survey.

Overall, the highest levels of occupational homogeneity are in states with low levels of unionization. This is also true for workers in the lowest-paid quintile of occupations. However, although the lowest levels of occupational homogeneity overall are in states with the highest levels of unionization, for workers in the lowest-paid quintile of occupations, the states with the highest levels of occupational homogeneity have middle levels of unionization.

Differences in occupational homogeneity trends between less and more unionized states (following equation (3)) show that establishments are growing more occupationally homogeneous over time in the less-unionized states, relative to the highly unionized states, by the predicted wage variance between-occupations. However, as shown in Table D1, there is no clear pattern of differences between more and less unionized states of trends in the normalized Herfindahl-Hirschman measure of occupational homogeneity

Following equation (4), the relationships between occupational homogeneity and wages are estimated separately for each unionization group of states, and these are shown in Table D2. Whether occupational homogeneity matters more for wages in more or less unionized states varies by the measure of occupational homogeneity, which workers are examined, and whether or not controls are included for establishment characteristics and occupation.

Table D1: Change in mean values of Occupational Homogeneity over time, by Unionization

Occupational Homogeneity Variable	Most Unionized States: Normalized Herfindahl of Occupational Homogeneity for the establishment	Most Unionized States: Partial Predicted Variance of Wages for the establishment	Least Unionized States: Normalized Herfindahl of Occupational Homogeneity for the establishment	Least Unionized States: Partial Predicted Variance of Wages for the establishment
All Occupations	**(32,864,786 observations)**		**(31,742,453 observations)**	
Raw Trend	0.01166	-0.00353	0.01118	-0.00740
	(0.00003)	(0.00001)	(0.00003)	(0.00001)
All Controls	0.00355	-0.00283	0.00564	-0.00712
	(0.00003)	(0.00001)	(0.00003)	(0.00001)
Lowest-paid quintile of occupations	**(2,610,856 observations)**		**(2,654,471 observations)**	
Raw Trend	0.01222	-0.00779	0.01772	-0.01241
	(0.00008)	(0.00002)	(0.00007)	(0.00002)
All Controls	0.00353	-0.00542	0.01110	-0.01008
	(0.00006)	(0.00002)	(0.00005)	(0.00002)
Second quintile of occupations	**(4,166,317 observations)**		**(4,003,654 observations)**	
Raw Trend	0.04135	-0.00888	0.02595	-0.00922
	(0.00008)	(0.00002)	(0.00008)	(0.00002)
All Controls	0.00570	-0.00338	0.00474	-0.00631
	(0.00006)	(0.00002)	(0.00006)	(0.00002)
Middle quintile of occupations	**(7,813,498 observations)**		**(7,508,503 observations)**	
Raw Trend	0.00744	-0.00036	0.01334	-0.00493
	(0.00006)	(0.00002)	(0.00006)	(0.00002)
All Controls	-0.00139	0.00131	0.00202	-0.00356
	(0.00005)	(0.00001)	(0.00005)	(0.00002)
Fourth quintile of occupations	**(6,983,356 observations)**		**(7,181,134 observations)**	
Raw Trend	-0.00294	-0.00030	0.00134	-0.00509
	(0.00008)	(0.00002)	(0.00007)	(0.00002)
All Controls	0.00012	-0.00142	0.00287	-0.00636
	(0.00006)	(0.00001)	(0.00006)	(0.00001)
Highest-paid quintile of occupations	**(11,290,759 observations)**		**(10,394,691 observations)**	
Raw Trend	0.00558	-0.00499	0.00391	-0.00922
	(0.00006)	(0.00002)	(0.00006)	(0.00002)
All Controls	0.00544	-0.00452	0.00399	-0.00862
	(0.00005)	(0.00002)	(0.00006)	(0.00002)

Note: These are coefficients α from regressions of the form $Occupation\ Homogeneity_{ijt} = \alpha Reference\ Date_t + \beta X_{ijt} + \varepsilon_{ijt}$, where the Reference Date is measured in decades since 2004 and X_{ijt} includes occupation fixed effects at the 6-digit SOC level, industry fixed effects at the 4-digit NAICS level, establishment size class, continuous establishment size, and state of location. Regressions are at the establishment-occupation-wage interval level, weighted by employment. Standard errors are in parentheses.

Table D2: Regressions of log wages on measures of Occupational Homogeneity, by Unionization group

	Most Unionized States		Least Unionized States	
Occupational Homogeneity Variable	Normalized Herfindahl of Occupational Homogeneity for the establishment	Partial Predicted Variance of Wages for the establishment	Normalized Herfindahl of Occupational Homogeneity for the establishment	Partial Predicted Variance of Wages for the establishment
All Occupations	**(32,864,786 observations)**		**(31,742,453 observations)**	
With only date fixed effects	-0.544	1.877	-0.517	1.805
	(0.000)	(0.001)	(0.000)	(0.001)
All Controls	-0.055	0.089	-0.065	0.099
	(0.000)	(0.000)	(0.000)	(0.000)
Lowest-paid quintile of occupations	**(2,610,856 observations)**		**(2,654,471 observations)**	
With only date fixed effects	-0.176	0.616	-0.168	0.588
	(0.000)	(0.001)	(0.000)	(0.001)
All Controls	-0.062	0.142	-0.079	0.168
	(0.000)	(0.001)	(0.000)	(0.001)
Second quintile of occupations	**(4,166,317 observations)**		**(4,003,654 observations)**	
With only date fixed effects	-0.222	0.762	-0.219	0.649
	(0.000)	(0.001)	(0.000)	(0.001)
All Controls	-0.083	0.341	-0.081	0.355
	(0.000)	(0.001)	(0.000)	(0.001)
Middle quintile of occupations	**(7,813,498 observations)**		**(7,508,503 observations)**	
With only date fixed effects	-0.144	0.516	-0.096	0.422
	(0.000)	(0.001)	(0.000)	(0.001)
All Controls	-0.077	0.315	-0.075	0.319
	(0.000)	(0.001)	(0.000)	(0.001)
Fourth quintile of occupations	**(6,983,356 observations)**		**(7,181,134 observations)**	
With only date fixed effects	-0.149	0.352	-0.247	0.733
	(0.000)	(0.001)	(0.000)	(0.001)
All Controls	-0.037	0.155	-0.072	0.218
	(0.000)	(0.001)	(0.000)	(0.001)
Highest-paid quintile of occupations	**(11,290,759 observations)**		**(10,394,691 observations)**	
With only date fixed effects	-0.115	0.204	-0.103	0.101
	(0.000)	(0.001)	(0.000)	(0.001)
All Controls	-0.009	-0.398	-0.025	-0.417
	(0.000)	(0.001)	(0.000)	(0.001)

Notes: These are coefficients α from regressions of the form $\ln(wage_{ijt}) = \alpha OccHomogeneity_{jt} + \beta X_{ijt} + \varepsilon_{ijt}$, where X includes reference date fixed effects, occupation fixed effects at the 6-digit SOC level, industry fixed effects at the 4-digit NAICS level, state fixed effects, and establishment size (using fixed effects for establishment size classes as well as continuous establishment size). All coefficients are significant at $p < 0.001$. Regressions are at the establishment-occupation-wage interval level, weighted by employment. Standard errors in parentheses.