NBER WORKING PAPER SERIES

INDUSTRIES, MEGA FIRMS, AND INCREASING INEQUALITY

John C. Haltiwanger
Henry R. Hyatt
James Spletzer

Working Paper 29920
http://www.nber.org/papers/w29920

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
April 2022, Revised December 2022

Any opinions and conclusions expressed herein are those of the authors and do not represent the views of the U.S. Census Bureau or the National Bureau of Economic Research. All results have been reviewed to ensure that no confidential information is disclosed (CBDRB-FY21-046, CBDRB-FY21-143, CBDRB-FY21-155, CBDRB-FY21-323, and CBDRB-FY22-356). We thank Keith Bailey, Nicholas Bloom, Andrew Foote, and participants at the 2021Wage Dynamics in the 21st Century conference, the 2021 Allied Social Science Associations meeting, the 2022 Washington Area Labor Economics Symposium, the 2022 Research Data Center conference, the 2022 Society of Labor Economists conference, the 2022 Society for Economic Dynamics conference, and 2022 NBER Summer Institute, and the 2022 Federal Reserve Bank of Atlanta Center for Human Capital Studies conference for helpful comments, and we especially thank our discussants Salomé Baslandze, Christina Patterson and Ayşegül Şahin. John Haltiwanger was also a Schedule A employee of the U.S. Census Bureau at the time of the writing of this paper.

Industries, Mega Firms, and Increasing Inequality
John C. Haltiwanger, Henry R. Hyatt, and James Spletzer
NBER Working Paper No. 29920
April 2022, Revised December 2022
JEL No. J21,J31

ABSTRACT

Most of the rise in overall earnings inequality is accounted for by rising between-industry dispersion from ten percent of 4-digit NAICS industries. These thirty industries are clustered especially in high-paying high-tech and low-paying retail sectors. The rise of employment in mega firms is concentrated in the industries that dominate rising earnings inequality. Earnings differentials for the mega firms relative to small firms decline in the low-paying industries but increase in the high-paying industries. A critical component accounting for the rising dispersion in the top thirty industries is an increasing covariance between industry premia and worker characteristics associated with high earnings.

John C. Haltiwanger
Department of Economics
University of Maryland
College Park, MD 20742
and NBER
haltiwan@econ.umd.edu

James Spletzer
Center for Economic Studies
U.S. Census Bureau
4600 Silver Hill Road
Washington, DC 20233
James.R.Spletzer@gmail.com

Henry R. Hyatt
Center for Economic Studies
U.S. Census Bureau
4600 Silver Hill Road
Washington, DC 20233
henry.r.hyatt@census.gov

1 Introduction

A growing number of studies attribute increases in earnings inequality to rising between-firm dispersion.[1] We confirm this pattern with comprehensive U.S. matched employer-employee data from 1996 to 2018. Our contribution is to explore and emphasize that rising between-firm dispersion mostly occurs at the industry level.[2] Rising between-industry dispersion accounts for most of the overall increase in earnings inequality, and is driven by a relatively small number of industries. About ten percent of 4-digit NAICS industries account for virtually all of the increase in between-industry dispersion, while accounting for less than 40% of employment. These industries are in the tails of the earnings distribution including high-paying industries such as Software Publishing (5112) and low-paying industries such as Restaurants and Other Eating Places (7225).[3] Remarkably, the remaining ninety percent of 4-digit industries individually contribute little to rising between-industry earnings inequality.

We provide further insights about rising between-industry inequality by quantifying the role of industry premia, between-industry sorting (high (low) wage workers increasingly employed in high (low) wage industries), and between-industry segregation (high (low) wage workers increasingly concentrated in specific industries). We follow Song et al. (2019) in using an Abowd, Kramarz, and Margolis (1999, hereafter AKM) decomposition of earnings as our intial method for quantifying these effects which facilitates the comparison of our results with those of Song et al. (2019). We also quantify these effects using a standard human capital equation (see Hoffman, Lee, and Lemieux (2020)) that relates earnings to age, education, occupation and industry effects. Finally, we also quantify these effects using the relationship between earnings with industry and occupation effects using the Occupational Establishment and Wage Survey (OEWS). Regardless of our approach, we find that

[1]Barth et al. (2016) and Song et al. (2019) provide evidence for the U.S. These papers follow an earlier literature emphasizing the importance of rising between-firm effects for earnings inequality that includes Davis and Haltiwanger (1991) and Dunne, Foster, and Haltiwanger (2004). Card, Heining, and Kline (2013) consider the role of firms in rising inequality in Germany, and Card, Cardoso, and Kline (2016) consider evidence from Portugal.

[2]Haltiwanger and Spletzer (2020a, 2020b) use a closely related data infrastructure and also emphasize the dominant contribution of rising between-industry dispersion. The first paper was a preliminary working paper that has been completely subsumed by the current paper. The second paper focuses on the changing patterns of labor market fluidity and the interaction with the changing structure of industries. The current paper is distinct in documenting and analyzing that a small fraction of industries account for virtually all the rising between-industry dispersion, that mega firms play a dominant role, and the decomposition of the between-industry contribution into sorting, segregation and industry premia. Briskar et al. (2020, 2022), in work produced concurrently with ours, have provided evidence for Italy on the dominant role of industry in firm-driven increases in inequality.

[3]Throughout this paper, we put 4-digit NAICS codes in parentheses.

an increase in sorting (i.e., an increasing covariance between industry premia and worker characteristics associated with high earnings) plays the most important role. Increased segregation of worker characteristics across industries is the second most important contributor. Rising between industry premia plays a supporting role.

The top ten percent of industries that contribute to rising inequality include nineteen that are high-paying. These industries account for 54.1% of the increase in between-industry inequality. The top three of these are high-paying, high-tech industries – Software Publishers (5112), Computer Systems Design (5415), and Other Information Services (5191) – and, in total, eleven of these nineteen high-paying industries are high-tech. As discussed in Oliner, Sichel, and Stiroh (2007) and Fernald (2014), these industries are characterized as the source of rapid technological advances. These industries play an outsized role in the tendency for high-paid workers to work both for high-paying firms (sorting) and with each other (segregation). Average worker earnings have surged in these industries while employment gains have been more modest.

Eleven low-paying industries are in the top ten percent of industries that dominate rising earnings inequality. These industries in combination account for 44.1% of the increase in between-industry inequality. More than one-fourth of the increase is accounted for by just three of these eleven: Restaurants and Other Eating Places (7225), Other General Merchandise Stores (4529), and Grocery Stores (4451). These industries have gone through substantial changes in recent decades, moving away from single establishment firms to large, national chains, see Foster, Haltiwanger and Krizan (2006), Foster et al. (2016), and Autor et al. (2020). While average earnings declined by a modest amount, substantial increases in employment in these low-paying industries led to especially large increases in inequality.

A distinctive feature of the dominant ten percent of industries is that they exhibit a sharp increase in the share of employment at mega firms, which we define as firms with more than 10,000 employees. Strikingly, the remaining ninety percent of industries exhibit small declines in the share of employment at mega firms. For the low-paying dominant industries, there is a sharp decline in the earnings of mega firms relative to earnings of the average industry (averaging over all 301 industries). This sharp decline is accompanied by a decline in the size-earnings premium within these low-paying industries. For the high-paying dominant industries, the mega firms experience a substantial increase in earnings relative to both small firms in the same industry and to earnings of the average industry. Thus, we find that the rise in "superstar" firms (see, e.g., Autor et al. (2020)) is concentrated in these dominant industries with accompanying systematic changes in the size-earnings premia.

Our findings build on the recent literature that highlights the dominant role of rising between-firm inequality. Our results are closest to those in the recent pathbreaking work of Song et al. (2019). Using Social Security Administration (SSA) administrative data linking employers and employees, they find a dominant role for rising between-firm earnings inequality. Moreover, using an AKM decomposition, they attribute most of this to changing composition from increasing sorting and segregation of workers across firms. Our analysis is based on using the comprehensive matched employer-employee data from the Longitudinal Employer-Household Dynamics (LEHD) data infrastructure at Census. While our results are consistent with the Song et al. (2019) findings, we depart significantly in that we highlight the dominant role of a small number of industries in accounting for rising between-firm inequality. In contrast, Song et al. (2019) do not report substantial differences across industry, although in a companion paper, Bloom et al. (2018) report an unusual missing data problem in their industry classifications.[4] Our results highlight that the increased sorting and segregation of workers across firms is across a relatively narrow set of industries. Our findings on between-industry sorting and segregation are also robust to using alternative approaches to AKM.

Our results contrast not only with the recent studies using SSA data but also with findings from household surveys such as the CPS. Haltiwanger, Hyatt, and Spletzer (2022) investigate this stark difference in findings from the CPS versus the findings in this paper generated from LEHD administrative matched employer-employee data. This companion paper finds that most (65.5%) of the rise in CPS earnings inequality is accounted for by rising between-industry inequality, which is remarkably similar to the 61.9% documented in this paper using administrative data. Estimating this 65.5% from the CPS required accounting for how workers are sorted into industries, linking the CPS and LEHD microdata, and replacing CPS industry sector codes with detailed LEHD 4-digit industry codes.

We shed further light on the role of industry by turning to the public domain OEWS that permits investigating the changing structure of occupations by detailed industry. We find that most of the increase in earnings dispersion in the OEWS data is accounted for by rising between-industry dispersion. Moreover, the top 30 industries we identified from the LEHD data account for about 96% of the rising between-industry dispersion in the OEWS data. About half of the increase in between-industry

[4]The industry analysis of Song et al. (2019) is reported on page 22 (first paragraph) and in Table 2 (page 17) of their paper. Bloom et al. (2018) report that industry codes in their SSA dataset were missing for all firms that entered after 2002. There are additional benefits to using the industry codes available through the LEHD, which we discuss below in Section 2.2.

dispersion is accounted for by increased sorting of high (low) paying occupations into high (low) paying industries. About one third is due to increased segregation of high and low paying occupations across the top 30 industries. The remainder is due to increased industry pay premia. There is also a close correspondence between changes in the occupational structure and our top 30 industries. Most of the increase in employment in top-paying occupations is accounted for by our 19 top-paying industries, while all of the increase in employment in low-paying occupations is in our 11 low-paying industries. For the top-paying industries, there is substantial increase in earnings differentials for the top-paying occupations. Likewise there is a substantial decline in earnings differentials for the low-paying occupations in our low-paying industries.

Our results imply that to explain rising inequality it is important to account for how the organization of production across firms within and especially between industries is changing. This perspective is well-recognized in the study of micro and macro productivity but has been neglected in the inequality literature which has focused on changes in the relative demand and supply of workers with different skills that are in turn have different abilities to accomplish different tasks. We don't view our results as inconsistent with that view but our findings highlight that the skill and task mix varies dramatically across firms and especially industries reflecting the different ways that production is organized. As such, when structural changes such as adoption of new technologies or globalization impact the relative demand of workers by skills and tasks this manifests itself through changes in the organization of production in distinct ways across detailed industries.

The paper proceeds as follows. Section 2 describes the data infrastructure and provides descriptive statistics about the changing distribution of earnings over our sample period from 1996 to 2018. Section 3 discusses the thirty industries that drive increasing inequality. The AKM decomposition methodology is presented in Section 4, followed by our extension to capture between-industry differences. Section 5 presents estimates of sorting, segregation, and firm premia from the AKM decomposition, distinguishing between- from within-industry differences. The rising importance of mega firms is discussed in Section 6. We explore how occupations have changed across industries in Section 7. Robustness of the contribution of industry premia, between-industry segregation and sorting to using a standard human capital decomposition of earnings is discussed in Section 8. Discussion of the interpretation of industry premia, between-industry sorting and between-industry segregation is provided in Section 9. Concluding remarks are in Section 10.

2 Data and descriptive statistics

2.1 The LEHD data and the analysis sample

We use Longitudinal Employer-Household Dynamics (LEHD) linked employer-employee data, which is created by the U.S. Census Bureau as part of the Local Employment Dynamics federal-state partnership. The LEHD data are derived from state-submitted Unemployment Insurance (UI) wage records and Quarterly Census of Employment and Wages (QCEW) data. Every quarter, employers who are subject to state UI laws – approximately 98% of all private sector employers, plus state and local governments – are required to submit to the states information on their workers (the wage records, which record the quarterly earnings of every worker in the firm) and their workplaces (the QCEW, which provides information on the industry and location of each establishment). The wage records and the QCEW data submitted by the states to the U.S. Census Bureau are enhanced with census and survey microdata in order to incorporate information about worker demographics (age, gender, and education) and the firm (firm age and firm size). Abowd et al. (2009) provide a thorough description of the source data and the methodology underlying the LEHD data. A job in the quarterly LEHD data is defined as the presence of an worker-employer match, and earnings is defined as the amount earned from that job during the quarter.

Because states have joined the LEHD program at different times, and have provided different amounts of historical data upon joining the LEHD program, the length of the time series of LEHD data varies by state. We use data from the 18 states that have data available from 1996:Q1 through 2018:Q4, which gives us annual data for 23 years.[5] We restrict the LEHD data to jobs in the private sector.

Following Song et al. (2019), we create annual person-level data from the quarterly job-level earnings data. We do this as $Y_t^i = \Sigma_j \Sigma_{q \in t} Y_{qt}^{ij}$, which sums the earnings Y that worker i receives from any firm j in any quarter q during year t. We use the Federal Employer Identification Number (EIN) as the firm identifier.[6] We follow Abowd, McKinney, and Zhao (2018) and delete any worker with 12 or more jobs during the year. A worker's employer in a given year is defined as the firm

[5]These 18 states are: CA, CO, CT, HI, ID, IL, KS, LA, MD, MN, MT, NC, NJ, OR, RI, TX, WA, and WI. These 18 states account for roughly 44% of national employment. The time series of employment from these 18 states closely tracks the national time series of total private sector employment published by the BLS.

[6]Haltiwanger and Spletzer (2020b) estimate variance decompositions using different levels of firm identifiers – the State UI account number, the EIN, and the enterprise. They find that rising between-industry dispersion accounts for most of the rising between-firm inequality regardless of the definition of the firm.

that contributes the worker's maximum earnings during the year. The annual data has 1.395 billion person-year observations (an average of about 61 million persons per year).

We create our analytical dataset following the sample restrictions of Song et al. (2019). We restrict to persons aged 20-60 and only keep person-year observations with annual real earnings $>$ \$3770 (=13 weeks * 40 hours per week * \$7.25 minimum wage), with nominal earnings converted to real terms using the 2013=100 PCE deflator. From this sample of 20-60 year olds with real annual earnings greater than \$3770, we topcode annual earnings at the 99.999% value (for anyone with earnings in the top 0.001%, we replace their earnings with the mean earnings of the top 0.001%). Our dataset has 1.048 billion person-year observations (an average of 45.6 million persons per year). All of our analysis uses real annual log earnings $y_t^i = \ln(Y_t^i)$. We define three 7-year intervals (1996-2002, 2004-2010, 2012-2018), reducing the sample to 959 million person-year observations. These 7-year intervals facilitate our variance decompositions.[7]

And finally, again following Song et al. (2019), we restrict to firm-year observations with 20 or more persons in the firm. This reduces our sample to 763 million person-year observations. Due to Census disclosure rules, we further restrict the firms with 20 or more employees in each year to have at least one male and one female; this means that the same set of firms can be used for separate variance decompositions by gender. The final LEHD data used to create all our results contains 758 million person-year observations (an average of about 36 million persons per year).

We also consider two other analytical samples as described below: one based on a CPS-LEHD linked sample and the other on the OEWS data.

2.2 Industry codes

Industry codes play a fundamental role in our analysis. We define industry at the 4-digit NAICS level.[8] Our basic results use establishment-level industry codes from the BLS QCEW program, aggregated to the Federal EIN level. Aggregation from establishment level data is done using maximum employment. If an EIN has $N > 1$ establishments with M industry codes, where $N \geq M > 1$, the

[7]The analysis in the main text uses pooled results for males and females. Separate results by gender are reported in Appendix C. Results are largely similar for females and males.

[8]The level of industry aggregation trades off tractability vs. comprehensiveness. Note that 4-digit NAICS industries aggregate 6-digit NAICS industries into "NAICS Industry Groups," which for ease of exposition, we refer to simply as "industries." Haltiwanger and Spletzer (2020b) measure rising inequality at different levels of NAICS aggregation, and demonstrate that the vast majority of rising between-industry inequality occurs at the 4-digit NAICS level.

industry code with the maximum employment is chosen for the aggregation.

Both BLS and Census have strong incentives and extensive statistical programs to assign detailed and accurate industry codes at the establishment-level. For BLS, the QCEW program yields high quality industry codes from the Annual Refiling Survey as well as the BLS surveys of businesses. For Census, the periodic surveys and the Economic Censuses of businesses provide rich sources of information on industry codes. BLS also shares their industry codes with Census. Census also obtains codes from SSA as part of the first step of identifying new businesses. The industry code from SSA is based on the information provided in the application for a new EIN (the SS-4 form). While SSA industry codes are a useful first step, Census has a clear hierarchy for industry codes in their Business Register and their business statistical programs, with the Economic Census (and related surveys) and BLS codes preferred (see Walker (1997)).

In complementary work, Haltiwanger and Spletzer (2020a) show that the fraction of the variance of earnings accounted for by industry effects is very similar using either BLS or Census codes but is much smaller using the industry codes Census obtains from SSA. Moreover, Bloom et al. (2018) indicate that the same SSA micro data used in Song et al. (2019) has missing industry codes for all new firms post 2002. Table 2 of Bloom et al. (2018, page 321) shows that EINs with missing industry codes increased from accounting for only 4% of total employment in 1980-1986 to 24% in 2007-2013 in their micro data. Our inference is that the high-quality industry codes from BLS and Census yield a more accurate characterization of the role of industry variation in accounting for earnings dispersion.

2.3 Descriptive statistics

Letting i index the worker, j the firm, k the industry, and t the year, we can write the variance of real annual log earnings y as:

$$\begin{aligned}
\underbrace{\mathrm{var}(y_t^{i,j,k,p} - \bar{y}^p)}_{\text{total dispersion}} &= \underbrace{\mathrm{var}(y_t^{i,j,k,p} - \bar{y}^{j,k,p})}_{\text{within-firm}} + \underbrace{\mathrm{var}(\bar{y}^{j,k,p} - \bar{y}^p)}_{\text{between-firm}} \\
&= \underbrace{\mathrm{var}(y_t^{i,j,k,p} - \bar{y}^{j,k,p})}_{\text{within-firm}} + \underbrace{\mathrm{var}(\bar{y}^{j,k,p} - \bar{y}^{k,p})}_{\substack{\text{between-firm,} \\ \text{within-industry}}} + \underbrace{\mathrm{var}(\bar{y}^{k,p} - \bar{y}^p)}_{\text{between-industry}}
\end{aligned} \tag{1}$$

We estimate this variance decomposition separately by 7-year intervals denoted by p. Note that $\bar{y}^p$ denotes average earnings among all workers in interval p, $\bar{y}^{j,k,p}$ the average earnings at firm j in

Table 1: Variance decomposition, by seven-year interval

	Interval 1: 1996-2002	Interval 2: 2004-2010	Interval 3: 2012-2018	Growth: 1 to 3
Variance, in levels:				
Total variance	0.794	0.862	0.915	0.121
Within-firm	0.512	0.532	0.531	0.018
Between-firm, within-industry	0.112	0.127	0.140	0.028
Between-industry	0.170	0.203	0.245	0.075
Variance, as percent of total:				
Within-firm	64.6%	61.7%	58.0%	14.9%
Between-firm, within-industry	14.0%	14.7%	15.3%	23.1%
Between-industry	21.4%	23.6%	26.8%	61.9%
Other measures:				
Sample size (millions)	239.4	249.2	269.7	
Number of firms (thousands)	470	460	466	
Number of NAICS industries	301	301	301	

Notes: Persons with annual real earnings > \$3770 in EINs with 20 or more employees. See Equation (1) for definitions.

interval p, and $\bar{y}^{k,p}$ the average earnings in industry k in interval p. Table 1 shows that for all workers, the variance of earnings increases from 0.794 in the first interval (1996-2002) to 0.915 in the third interval (2012-2018).[9] Of this 0.121 increase, 0.018 occurs within firms (14.9%), 0.028 between firms but within industries (23.1%), and 0.075 between industries (61.9%). These estimates state that between-industry variance growth accounts for 72.8% $(= 0.075/(0.028 + 0.075))$ of the between-firm contribution to increasing inequality.

It is important to distinguish between a cross-sectional variance decomposition versus a growth decomposition. At a given point in time, the majority of variance is within firms: 64.6% of variance in the first interval is within firms, 61.7% in the second interval, and 58.0% in the third interval. This declining relative percentage is indicative that the within-firm person component of earnings variance is becoming less important over time. Growth in the within-industry firm component is positive but much smaller than between-industry growth. It is the between-industry component that is growing substantially over time, from 21.4% in the first interval to 26.8% in the third interval.

[9] We follow Song et al. (2019) in using 7-year intervals which facilitates the estimation of the AKM earnings decomposition for different time intervals. Appendix Figure A2 shows an annual version of this decomposition. Appendix Figure A1 reports related basic facts using the annual version of our analytical sample.

2.4 Earnings percentiles

The statistics in Table 1 demonstrate that 4-digit NAICS industry accounts for almost two-thirds of the growth of earnings variance. In this section, we present a descriptive analysis to learn where in the earnings distribution industry is important. We first estimate annual earnings for each of the percentiles 1 to 99 for the first (1996-2002) and the third (2012-2018) 7-year intervals, and then calculate the difference between the first and third intervals for each percentile, as shown in Figure 1.[10] In our analytical sample, comparing the first and the third intervals, annual earnings declined by more than five log points for the first 34 percentiles, and declined for the first 61 percentiles. However, earnings at the top increased substantially. Earnings in the top 23 percentiles increased by more than 5 log points (5.1%), and earnings in the top 13 percentiles increased by more than 10 log points (10.5%).[11]

We use a simple decomposition to understand how the person, the firm, and the industry help account for the changing distribution of earnings. We can express the difference between earnings $y_t^{i,j,k,p}$ and average earnings $\bar{y}^p$ as

$$\underbrace{y_t^{i,j,k,p}-\bar{y}^p}_{\text{relative earnings}} = \underbrace{y_t^{i,j,k,p}-\bar{y}^{j,k,p}}_{\text{within-firm}} + \underbrace{\bar{y}^{j,k,p}-\bar{y}^{k,p}}_{\substack{\text{between-firm,}\\ \text{within-industry}}} + \underbrace{\bar{y}^{k,p}-\bar{y}^p}_{\text{between-industry}} \tag{2}$$

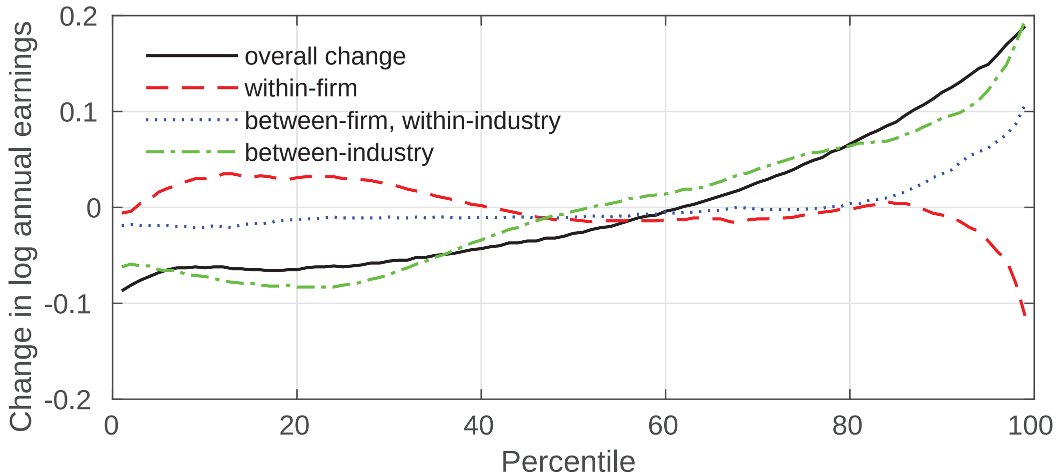
We estimate the mean of each of the terms on the right-hand side for each percentile of relative worker earnings ($y_t^{i,j,k,p}-\bar{y}^p$), noting that firm mean earnings $\bar{y}^{j,k,p}$, industry mean earnings $\bar{y}^{k,p}$, and the grand mean of earnings $\bar{y}^p$ are from the full sample of workers rather than calculated within each percentile. To interpret this exercise, think of workers in the first percentile, who have earnings between the ½th and 1½th percentiles. We estimate how the earnings of these workers differ from the earnings of their firm ($y_t^{i,j,k,p}-\bar{y}^{j,k,p}$), how the earnings of their firm differ from the earnings of their industry ($\bar{y}^{j,k,p}-\bar{y}^{k,p}$), and how the earnings of their industry differ from the grand mean of earnings ($\bar{y}^{k,p}-\bar{y}^p$). We do this for each percentile in the first and third intervals, and then calculate the difference between the first and third intervals for each percentile.

For each percentile, the dashed line in Figure 1 is the person component $y_t^{i,j,k,p}-\bar{y}^{j,k,p}$, the dotted

[10]For each 7-year interval p, we create percentiles $x \in \{1,2,\ldots,99\}$ for $y_t^{i,j,k,p}-\bar{y}^p$ where percentile X is defined as the mean of $y_t^{i,j,k,p}-\bar{y}^p$ for all workers between the $x-1/2$ and the $x+1/2$ percentiles.

[11]Throughout this paper, we convert any log differential x into a proportionate change using the expression e^x-1. For small differences, log points (i.e., log differentials multiplied by 100) are approximately equal to the percent change.

Figure 1: Change in log real annual earnings, by percentile



Notes: Persons with annual real earnings > \$3770 in EINs with 20 or more employees. See Equation (2) for definitions.

line is the firm component $\bar{y}^{j,k,p} - \bar{y}^{k,p}$, and the dash-dot line is the industry component $\bar{y}^{k,p} - \bar{y}^{p}$. We see that at the lower end of the earnings distribution, industry accounts for most of the decline. At the higher end of the earnings distribution, industry also plays a sizeable role in accounting for increasing earnings. Looking ahead to our subsequent results, Figure 1 suggests that industry plays a major role in understanding earnings change at both the lower and the upper ends of the earnings distribution.

Of interest is the role of the between-firm, within-industry component in Figure 1. This component $\bar{y}^{j,k,p} - \bar{y}^{k,p}$ has only a modest contribution to the changing earnings distribution for the first 87 percentiles. The absolute value of the dotted line is less than 2.5 log points (2.5%) for each of the first 87 percentiles. From the 88th to the 99th percentiles, the between-firm, within-industry component increases monotonically to a value of 10.7 log points (11.3%) for the highest percentile.

3 The industries that drive increasing inequality

3.1 The top ten percent of industries vs. the remaining ninety percent

We have demonstrated that almost two-thirds of the growth in inequality occurs between rather than within industries. We now propose a measure of a particular industry's contribution to inequality and assess how this varies across industries. Formally, consider the total between-industry contribution to

Table 2: Industry contributions to between-industry variance growth, by variance contribution

Industry share of between-industry variance growth	Number of industries	Total employment share	Total contribution to between-industry variance growth	Total share of between-industry variance growth
$> 5\%$	5 industries	8.8%	0.031	40.7%
1% to 5%	25 industries	30.5%	0.043	57.4%
0.05% to 1%	71 industries	21.8%	0.017	22.3%
-0.05% to 0.05%	145 industries	19.3%	-0.000	-0.1%
$< -0.05\%$	55 industries	19.7%	-0.015	-20.3%
Overall	301 industries	100.0%	0.075	100.0%

Notes: Persons with annual real earnings > \$3770 in EINs with 20 or more employees. Employment shares are calculated as the average of 1996-2002 and 2012-2018 employment shares. See Equation (3) for definitions.

inequality, which is given by $\mathrm{var}(\bar{y}^{k,p} - \bar{y}^{p})$. Between-industry variance growth is then

$$\underbrace{\Delta \mathrm{var}(\bar{y}^{k,p} - \bar{y}^{p})}_{\substack{\text{between-industry} \\ \text{variance growth}}} = \underbrace{\Sigma_{k=1}^{301} \Delta \underbrace{\left(\frac{N^{k,p}}{N^{p}}\right)}_{\substack{\text{employment} \\ \text{share}}} \underbrace{(\bar{y}^{k,p} - \bar{y}^{p})^2}_{\substack{\text{relative} \\ \text{earnings}}}}_{\substack{\text{industry } k\text{'s contribution} \\ \text{to between-industry} \\ \text{variance growth}}}, \tag{3}$$

where N counts worker-employer-year combinations (i.e., employment), $N^{k,p}$ is total employment in industry k in interval p, and N^{p} is total employment in interval p. We define industry k's contribution to between-industry variance growth as $\Delta(\frac{N^{k,p}}{N^{p}})(\bar{y}^{k,p} - \bar{y}^{p})^2$.

There are a total of 301 4-digit NAICS industries in our LEHD data. A natural starting point is to group industries by their contributions to increasing inequality, which we explore in Table 2. There are five industries that each contribute more than 5% of between-industry variance growth, accounting for 40.7% of between-industry variance growth. These five industries have 8.8% of total employment. An additional twenty-five industries each contribute between 1% and 5% of between-industry variance growth, accounting for 57.4% of between-industry variance growth. In total, the top thirty industries – about ten percent of all 4-digit NAICS industries – account for 98.1% of between-industry variance growth and 39.3% of employment.

As nearly two-thirds of the growth in U.S. earnings dispersion has occurred between industries rather than within them, these thirty industries account for most of increasing inequality. We provide

Table 3: Industry contributions to between-industry variance growth, top 30 industries

4-digit NAICS	Industry title	Employment share: average	Employment share: change	Relative earnings: average	Relative earnings: change	Share of bet.-ind. variance growth
2111	Oil & Gas Extraction	0.3%	-0.1%	1.012	0.247	1.8%
2131	Support Activities for Mining	0.5%	0.3%	0.374	0.191	1.4%
3254	Pharmaceutical Manufacturing	0.5%	-0.1%	0.799	0.203	1.6%
3344	Semiconductor Manufacturing	0.8%	-0.5%	0.556	0.299	1.4%
4234	Professional Equip. Wholesaler	0.7%	-0.0%	0.557	0.190	1.9%
4441	Building Material & Supplies	0.9%	0.1%	-0.293	-0.180	1.5%
4451	Grocery Stores	2.3%	0.0%	-0.378	-0.194	4.7%
4481	Clothing Stores	0.7%	-0.0%	-0.607	-0.244	2.6%
4529	Othr. Genrl. Merchandise Stores	1.4%	1.5%	-0.539	-0.051	6.8%
5112	Software Publishers	0.5%	0.2%	1.009	0.186	5.6%
5182	Data Processing Services	0.3%	-0.0%	0.545	0.301	1.3%
5191	Other Information Services	0.2%	0.3%	0.798	0.699	5.8%
5221	Depository Credit Intermediat.	2.1%	0.0%	0.189	0.234	2.5%
5231	Securities Brokerage	0.5%	-0.1%	0.866	0.204	1.1%
5239	Other Financial Invest. Activity	0.3%	0.1%	0.834	0.388	3.3%
5241	Insurance Carriers	1.6%	-0.4%	0.488	0.167	2.3%
5413	Architectur. & Enginr. Services	1.2%	0.1%	0.469	0.161	2.6%
5415	Computer Systems Design	1.7%	0.9%	0.663	0.012	5.6%
5416	Management & Scientific Serv.	0.9%	0.6%	0.381	0.069	1.8%
5417	Scientific Research Services	0.8%	-0.1%	0.741	0.244	3.3%
5511	Management of Companies	2.0%	-0.1%	0.471	0.201	5.0%
5613	Employment Services	3.9%	0.6%	-0.685	0.017	2.5%
5617	Services to Buildings & Dwell.	1.1%	0.3%	-0.493	-0.002	1.1%
6211	Offices of Physicians	1.7%	0.5%	0.254	0.099	1.6%
6216	Home Health Care Services	0.8%	0.4%	-0.525	-0.016	1.7%
6221	General Medical & Hospitals	4.5%	0.5%	0.205	0.162	4.2%
6233	Continuing Care Retirement	0.6%	0.4%	-0.493	-0.001	1.2%
6241	Individual & Family Services	0.8%	0.6%	-0.490	-0.155	3.5%
7139	Othr. Amusement & Recreation	0.6%	0.1%	-0.594	-0.106	1.7%
7225	Restaurants & Othr. Eat Places	4.9%	2.0%	-0.739	-0.027	16.9%

Notes: Persons with annual real earnings $>$ \$3770 in EINs with 20 or more employees. Average log earnings for industry k are relative to the economy average. The 1996-2002 and 2012-2018 intervals are averaged. Changes are the growth (or decline) from 1996-2002 to 2012-2018. See Equation (3) for definitions.

Table 4: Industry contributions to between-industry variance growth, by average earnings

Industry relative earnings	Number of industries	Total employ-ment share	Total contribution to bet.-ind. var. growth	Total share of of bet.-ind. var. growth	Shift-share: employment	earnings
		30 industries with variance contribution $> 1\%$				
High-paying	19 industries	21.1%	0.041	54.1%	16.1%	83.9%
Low-paying	11 industries	18.1%	0.033	44.1%	68.3%	31.7%
		271 industries with variance contribution $\leq 1\%$				
High-paying	146 industries	34.9%	0.001	1.3%		
Low-paying	125 industries	25.9%	0.000	0.6%		
Overall	301 industries	100.0%	0.075	100.0%	14.0%	86.0%

Notes: Persons with annual real earnings > \$3770 in EINs with 20 or more employees. Employment shares are calculated as the average of 1996-2002 and 2012-2018 employment shares. Industry k's contribution to between-industry variance growth is specified in Equation (3). The shift-share calculations for changing employment and earnings follow Equation (4). Shift-share results are summed across industries and normalized by the total contribution so that the two components sum to 100%. The two rows for the 271 industries with variance contribution $\leq 1\%$ have missing cells because the denominator for the shift-share decomposition is close to zero.

detail about these thirty industries in Table 3 (the industries in Table 3 are sorted by NAICS). The largest contribution is from Restaurants and Other Eating Places (7225), which alone accounts for 16.9% of between-industry variance growth. The second-largest contribution occurs among Other General Merchandise Stores (4529), which accounts for 6.8%. While the most important two industries to increasing inequality tend to offer low-paying jobs, the other three industries that account for more than 5% of between-industry variance growth are high-paying: Software Publishers (5112), Computer Systems Design (5415), and Management of Companies (5511).

What about the other 271 4-digit NAICS industries? The contributions of these industries to between-industry variance growth are summarized in Table 2. There are 145 industries that each contribute approximately 0.0% (to be precise, greater than -0.05% and less than 0.05%) to between-industry variance growth. This says that almost half of all 4-digit NAICS industries contribute essentially nothing to inequality growth. There are 71 industries that contribute between 0.05% and 1.0%, accounting for 22.3% of between-industry variance growth. These industries are basically offset by another 55 industries that have a negative contribution ($< -0.05\%$), accounting for -20.3% of between-industry variance growth.

As seen in Table 4, the top thirty industries include nineteen high-paying industries that account for 54.1% of between-industry variance growth, and eleven low-paying industries that account for 44.1% of between-industry variance growth. The other 271 industries that have small contributing and offsetting contributions to increasing inequality do not occur systematically among high-paying vs. low-paying industries. 146 high-paying industries account for 1.3% of between-industry variance growth, and 125 low-paying industries account for only 0.6% of between-industry variance growth.

Changes in earnings and employment share determine an industry's contribution to growth in inequality. This is seen in the expression defining industry k's contribution to between-industry variance growth: $\Delta(\frac{N^{k,p}}{N^p})(\bar{y}^{k,p}-\bar{y}^p)^2$. If an industry with relatively high earnings exhibits an earnings increase, then, *ceteris paribus*, inequality will increase. Analogously, inequality will increase if an industry with relatively low earnings exhibits an earnings decrease, *ceteris paribus*. Note that changes in earnings towards the mean will tend to reduce inequality. For example, if the earnings of a high-paying industry moves closer to the average among all workers, this decline would reduce earnings inequality.

Employment shares also determine industry-level contributions to inequality. An industry's earnings changes will have larger effects on inequality when its employment share is larger. Changes in an industry's employment share will have smaller effects on inequality when that industry's pay is more similar to the overall average. Employment gains among very high- and very low-paying industries tend to increase inequality.

In Table 4, we report the relative importance of earnings changes vs. employment changes using a shift share decomposition. Industry k's contribution to between-industry variance growth is $\Delta(\frac{N^{k,p}}{N^p})(\bar{y}^{k,p}-\bar{y}^p)^2$. We can use a standard shift-share decomposition to express this change in terms of the components attributable to changes in employment vs. earnings:

$$\underbrace{\Delta(\frac{N^{k,p}}{N^p})(\bar{y}^{k,p}-\bar{y}^p)^2}_{\substack{\text{industry } k\text{'s contribution}\\ \text{to between-industry}\\ \text{variance growth}}} = \underbrace{\overline{(\bar{y}^{k,p}-\bar{y}^p)^2}\Delta(\frac{N^{k,p}}{N^p})]}_{\text{shift-share: employment}} + \underbrace{\frac{\overline{N^{k,p}}}{N^p}\Delta(\bar{y}^{k,p}-\bar{y}^p)^2}_{\text{shift-share: earnings}}, \tag{4}$$

where $\overline{(\bar{y}^{k,p}-\bar{y}^p)^2}$ and $\overline{\frac{N^{k,p}}{N^p}}$ are averages of intervals 1 and 3. We do this for our top thirty industries, distinguished by high-paying and low-paying industries (we do not present the shift share estimates for the other 271 industries since the denominator of the shift share is very close to zero). Among the nineteen high-paying industries, 83.9% of between-industry variance growth is accounted for by changing relative earnings, and the remaining 16.1% is accounted for by changing employment shares. Among the eleven low-paying industries, the relative importance of earnings vs. employment is reversed: 68.3% of between-industry variance growth is accounted for by changing employment shares, and the remaining 31.7% is accounted for by changing relative earnings. These results highlight different explanations for why between-industry variance growth is increasing at the opposite tails of the earnings distribution. Inequality growth at the top of the earnings distribution is a story of increasing earnings, whereas inequality growth at the bottom of the earnings distribution is a story of increasing employment.

These two different explanations for increasing inequality among low- vs. high-paying industries is evident in the earnings and employment changes of the thirty industries listed in Table 3. All of the nineteen high-paying industries exhibit earnings increases during our time period. The most rapid growth is found in Other Information Services (5191), which had a 69.9 log point (101.2%) increase in relative earnings.[12] Of the remaining high-paying industries, nine had earnings increases in excess of 20 log points (22.1%), six had increases between 10 (10.5%) and 20 log points, and three had increases less than 10 log points.

Most of the eleven low-paying industries exhibit earnings decreases, yet they are smaller in absolute value than the earnings increases among the high-paying industries. The only low-paying industry with a decline greater in magnitude than 20 log points (22.1%) is Clothing Stores (4481), which had a 24.4 log point (27.6%) decrease in relative earnings. Of the remaining low-paying industries, four had earnings declines between 10 (10.5%) and 20 log points, five had earnings declines between 0 and 10 log points. One industry, Employment Services (5613), exhibited a relatively small increase in earnings.

On the other hand, changes in employment are more important for the eleven low-paying industries than for the nineteen high-paying industries. Two low-paying industries in Table 3 stand out: Restaurants and Other Eating Places (7225) had a 2.0 percentage point increase in employment share, and Other General Merchandise Stores (4529) had a 1.5 percentage point increase in employment share. Eight of the other low-paying industries have smaller employment share increases (less than one percentage point), and one industry (Clothing Stores, 4481) had a declining employment share. Among the nineteen high-paying industries, none had employment share increases exceeding one percentage point, ten had small employment share increases (less than one percentage point), and about

[12] We convert log differentials to proportionate changes using the expression $e^x - 1$. For small differences, log points are approximately equal to the percentage change.

half (nine) of the high-paying industries had declining employment shares.

3.2 Characteristics of the top thirty industries

Are there common characteristics that underlie the top ten percent of industries that contribute to between-industry variance growth? The top thirty industries reflect a small number of industry clusters that are notable for undergoing structural transformations that have been the subject of independent analysis. Eleven of the nineteen high-paying industries have been defined as high-tech industries in terms of STEM intensity by Hecker (2005) and Goldschlag and Miranda (2016). These innovative industries in combination account for about one-third of the between-industry increase in earnings dispersion. The transformation of the retail sector accounts for another one-third of the increase.[13]

Other industry clusters evident in Table 3 include four of the nine 4-digit industries in Finance and Insurance (NAICS sector 52), Management of Companies (NAICS sector 55), two of the eleven 4-digit industries in Administrative and Support Services (NAICS sector 56, e.g., Employment Services (5613)), and two of the five 4-digit industries in Mining (NAICS sector 21, e.g., Oil and Gas Extraction (2111)). Finance and Insurance (NAICS sector 55) industries have undergone tremendous restructuring and consolidation following deregulation (see, e.g., Kroszner and Strahan (2014)). Management occupation differentials have risen dramatically over our sample period (see Haltiwanger and Spletzer (2020a)). The Employment Services industry (5613) is a low-paying industry that has experienced dramatic growth and change with the growth of Professional Employee Organizations (NAICS 561330, see Dey et al. (2006)). Oil and Gas Extraction (2111) has long been a high-paying industry and underwent dramatic expansion and innovation with the shale oil boom starting in 2007 (Decker et al. (2016)).

The changing structure of businesses in Health Care and Social Assistance (NAICS sector 62) also plays an important role with four of the thirty top industries.[14] In combination, these industries account for 12.2% of the increase in between-industry inequality. In this sector, both high-paying and low-paying industries are important contributors. High-paying industries such as Offices of Physicians (6211) and General Medical and Surgical Hospitals (6221) contribute substantially, as well as low-paying industries such as Home Health Care (6216) and Retirement Care Facilities (6233). In-

[13] These include the large industries in retail that have undergone the most significant business model transformations towards large national chains (see Foster et al. (2006, 2016)).

[14] This reflects four out of eighteen 4-digit industries in the Health Care and Social Assistance (NAICS sector 62).

creased consolidation of hospitals and physician offices has been the subject of active research (see, e.g., Fulton (2017) and Cooper et al. (2019)). Much of the focus of that literature has been on rising concentration and markups. We find that these industries are high-paying industries with increases in both relative earnings and employment shares.

Only two of the 86 4-digit Manufacturing industries (NAICS sector 31-33) are among our top thirty industries – Pharmaceuticals (3254) and Semiconductors (3344), and both are part of the high-tech, STEM-intensive eleven. Manufacturing industries, on average, have higher than average earnings compared to other industries with a modest increase in relative earnings over time. However, most high-paying high-tech manufacturing industries have exhibited declines in employment share mitigating the impact of earnings increases on rising inequality. Indeed, Computer Manufacturing (3341) is the industry with the largest negative contribution (-1.6%, as shown in Appendix Table A4) to between-industry variance growth.

4 Empirical framework

4.1 Decomposing Earnings using AKM

To understand the role of workers and firms in the generation of earnings inequality, we begin by using the linear model of AKM to make our results directly comparable with Song et al. (2019). We estimate our model separately for each of three seven-year intervals: 1996-2002, 2004-2010, and 2012-2018. Following Song et al. (2019), we assume that earnings $y_t^{i,j,k,p}$ are the sum of the effect $\theta^{i,p}$ of worker i in interval p, a firm effect $\psi^{j,k,p}$ when employed by employer j in industry k during interval p, and a vector of time-varying observable characteristics $X_t^{i,p}$ for worker i at time t, which have distinct marginal effects β^p by interval p. We can express this as

$$y_t^{i,j,k,p} = X_t^{i,p}\beta^p + \theta^{i,p} + \psi^{j,k,p} + \varepsilon_t^{i,j,k,p}. \tag{5}$$

Our observable characteristics control for time and worker age. Specifically, we include a set of year dummies that capture calendar year effects on earnings. To control for worker age, we follow the specification of Card, Cardoso, and Kline (2016). We center age around 40, include a quadratic and cubic transformation of worker age, but omit the linear term. To solve this model, we implement the iterative method proposed by Guimarães and Portugal (2010).[15]

The AKM approach to decomposing earnings has received substantial scrutiny in terms of the interpretation of the estimated person and firm effects. Recent research has highlighted (see Bonhomme et al. (2022)) the potential for *limited mobility bias* arising from the small number of transitions per firm.[16] They find that limited mobility bias yields an upward bias in the variance of firm effects and a downward bias in the covariance between firm and worker effects. However, they find little bias on the contribution of the change in the role of premia and sorting for the change in inequality (and it is the latter that is the focus of our work). We use AKM to make our results comparable to Song et al. (2019) which enables us to highlight that the sorting and segregation effects they identified are mostly occurring at the industry level. However, we also include (see Sections 7 and 8) alternative decompositions based on the OEWS data and the CPS-LEHD linked data.

4.2 Sorting, segregation, and pay premia

The AKM framework allows for a rich decomposition of earnings dispersion. It is possible to express the variance of earnings in terms of the dispersion of worker and firm effects, the effects of observable characteristics, their covariances, and the dispersion of the residual. Following Song et al. (2019), denote the firm-level average worker effect of firm j during interval p (hereafter suppressing the superscript for interval p) as $\bar{\theta}^{j,k}$, and similarly denote the average observable characteristics as $\bar{X}^{j,k}$. The variance of earnings can be written as

$$
\begin{aligned}
\operatorname{var}(y_t^{i,j,k}) = & \underbrace{\operatorname{var}(\theta^i - \bar{\theta}^{j,k}) + \operatorname{var}(X_t^i\beta - \bar{X}^{j,k}\beta) + 2\operatorname{cov}(\theta^i - \bar{\theta}^{j,k}, X_t^i\beta - \bar{X}^{j,k}\beta)}_{\text{within-firm person effects and observables}} + \\
& \underbrace{\operatorname{var}(\bar{\theta}^{j,k}) + \operatorname{var}(\bar{X}^{j,k}\beta) + 2\operatorname{cov}(\bar{\theta}^{j,k}, \bar{X}^{j,k}\beta)}_{\text{total segregation}} + \underbrace{\operatorname{var}(\psi^{j,k})}_{\text{total pay premia}} + \\
& \underbrace{2\operatorname{cov}(\bar{\theta}^{j,k}, \psi^{j,k}) + 2\operatorname{cov}(\bar{X}^{j,k}\beta, \psi^{j,k})}_{\text{total sorting}} + \underbrace{\operatorname{var}(\varepsilon_t^{i,j,k})}_{\text{residual (within-firm)}} .
\end{aligned}
\tag{6}
$$

[15]We also estimated the AKM decomposition separately for females and males. We find that qualitatively and quantitatively the AKM decomposition results are similar for females and males. To facilitate comparisons with Song et al. (2019) who focus on results for males in the main text of their paper and report results for females in an appendix, we report the results for females and males separately in Appendix C.

[16]There has also been concern raised about exogenous mobility. Bonhomme et al. (2022) have highlighted that this is less of an issue than limited mobility bias. The reason is that what is required is that mobility is unrelated to the residual from the AKM model after controlling for person and firm effects.

The worker- and firm-level contributions can be collected from the terms of the basic decomposition.

Between-firm dispersion can be expressed through the contributions of sorting, segregation, and firm premia. Sorting is the covariance between worker and firm effects, given by $2\text{cov}(\bar{\theta}^{j,k}, \psi^{j,k}) + 2\text{cov}(\bar{X}^{j,k}\beta, \psi^{j,k})$. In other words, sorting reflects the extent to which highly-paid workers work for high-paying firms. Segregation reflects the concentration within firms of workers of the same type (captured by person effects), given by $\text{var}(\bar{\theta}^{j,k}) + \text{var}(\bar{X}^{j,k}\beta) + 2\text{cov}(\bar{\theta}^{j,k}, \bar{X}^{j,k}\beta)$. The remaining contributor to between-firm dispersion is reflected in the firm premia term $\text{var}(\psi^{j,k})$.

The remaining dispersion is within-firm dispersion. Worker-level effects are given by $\text{var}(\theta^i - \bar{\theta}^{j,k}) + \text{var}(X_t^i\beta - \bar{X}^{j,k}\beta) + 2\text{cov}(\theta^i - \bar{\theta}^{j,k}, X_t^i\beta - \bar{X}^{j,k}\beta)$. Residual dispersion $\text{var}(\varepsilon_t^{i,j,k})$ occurs within firms.

4.3 Industry-enhanced variance decomposition

We now propose a tractable framework for the study of inequality in terms of effects that occur within- and between-industries. To explore cross-industry differences, we calculate industry-level averages. We define the average worker effect in industry k in interval p as $\bar{\theta}^k$, the average effect of observable characteristics as $\bar{X}^k\beta$, and the average firm effect as $\bar{\psi}^k$.[17] Given this notation, it is possible to distinguish between how firm-level pay premia relate to within- vs. between-industry earnings dispersion. This is given by

$$
\begin{aligned}
\text{var}(y_t^{i,j,k}) = {} & \underbrace{\text{var}(\theta^i - \bar{\theta}^{j,k}) + \text{var}(X_t^i\beta - \bar{X}^{j,k}\beta) + 2\text{cov}(\theta^i - \bar{\theta}^{j,k}, X_t^i\beta - \bar{X}^{j,k}\beta)}_{\text{within-firm person effect and observables}} + \\
& \underbrace{\text{var}(\bar{\psi}^k)}_{\substack{\text{between-industry}\\ \text{pay premia}}} + \underbrace{\text{var}(\psi^{j,k} - \bar{\psi}^k)}_{\substack{\text{within-industry,}\\ \text{between-firm}\\ \text{pay premia}}} + \underbrace{2\text{cov}(\bar{\theta}^k, \bar{\psi}^k) + 2\text{cov}(\bar{\psi}^k, \bar{X}^k\beta)}_{\text{between-industry sorting}} + \\
& \underbrace{2\text{cov}[(\bar{\theta}^{j,k} - \bar{\theta}^k), (\psi^{j,k} - \bar{\psi}^k)] + 2\text{cov}[(\psi^{j,k} - \bar{\psi}^k), (\bar{X}^{j,k}\beta - \bar{X}^k\beta)]}_{\text{within-industry, between-firm sorting}} + \\
& \underbrace{\text{var}(\bar{\theta}^{j,k} - \bar{\theta}^k) + \text{var}(\bar{X}^{j,k}\beta - \bar{X}^k\beta) + 2\text{cov}[(\bar{\theta}^{j,k} - \bar{\theta}^k), (\bar{X}^{j,k}\beta - \bar{X}^k\beta)]}_{\text{within-industry, between-firm segregation}} + \\
& \underbrace{\text{var}(\bar{\theta}^k) + \text{var}(\bar{X}^k\beta) + 2\text{cov}(\bar{\theta}^k, \bar{X}^k\beta)}_{\text{between-industry segregation}} + \underbrace{\text{var}(\varepsilon_t^{i,j,k})}_{\text{residual (within-firm)}}
\end{aligned}
\tag{7}
$$

[17] This approach to measuring industry-level pay premia has recently been used by Card, Rothstein, and Yi (2022) in their analysis of between-industry pay differentials.

The notation is somewhat more complicated than Section 4.2 given the definitions we start with, but the intuition is analogous. $\mathrm{var}(\psi^{j,k}) = \mathrm{var}(\bar{\psi}^k) + \mathrm{var}(\psi^{j,k} - \bar{\psi}^k)$, where $\bar{\psi}^k$ reflects the between-industry dispersion in average firm effects, i.e. industry-level pay premia. The remaining term $\mathrm{var}(\psi^{j,k} - \bar{\psi}^k)$ captures the within-industry dispersion of firm-level pay premia. In addition to pay premia, we can distinguish between the within- vs. between-industry components of sorting and segregation.

Between-industry sorting is defined as $2\mathrm{cov}(\bar{\theta}^k, \bar{\psi}^k) + 2\mathrm{cov}(\bar{\psi}^k, \bar{X}^k\beta)$. It therefore reflects the extent to which highly-paid workers are employed in industries with a high pay premium. This is distinct from within-industry sorting $2\mathrm{cov}[(\bar{\theta}^{j,k} - \bar{\theta}^k), (\theta^{j,k} - \bar{\theta}^k)] + 2\mathrm{cov}[(\theta^{j,k} - \bar{\theta}^k), (\bar{X}^{j,k}\beta - \bar{X}^k\beta)]$ This is the component of sorting where relatively highly-paid workers tend to work at high-paying firms, apart from industry-level differences. For example, workers and firms in Restaurants and Other Eating Places (7225) industry may tend to have low worker effects, while those among Software Publishers (5112) may have high effects. The between-industry component reflects these industry level differences. The within-industry component reflects the extent to which relatively low- vs. high-paid workers work for relatively low-vs. high-paying firms in those industries.

Segregation also can be decomposed into its within- vs. between-industry components. Between-industry segregation is given by industry-level average worker effects. Formally, this is expressed as $\mathrm{var}(\bar{\theta}^k) + \mathrm{var}(\bar{X}^k\beta) + 2\mathrm{cov}(\bar{\theta}^k, \bar{X}^k\beta\ )$. This is the extent to which low- vs. highly-paid workers tend to work with each other. To continue with the previous example, Restaurants and Other Eating Places (7225) may employ workers with a low person effect, on average, while employers among Software Publishers (5112) may employ workers with a high average person effect. The extent to which this is related to the firm-level pay differences reflects sorting. The extent to which it reflects similar workers grouped together is segregation. Segregation that occurs within industries is expressed as $\mathrm{var}(\bar{\theta}^{j,k} - \bar{\theta}^k) + \mathrm{var}(\bar{X}^{j,k}\beta - \bar{X}^k\beta) + 2\mathrm{cov}[(\bar{\theta}^{j,k} - \bar{\theta}^k), (\bar{X}^{j,k}\beta - \bar{X}^k\beta)]$.

We have now defined the within- vs. between-industry contributions of sorting, segregation, and pay premia to inequality. Observe that this only required further decomposition of between-firm inequality. Within-firm, worker-level dispersion, as well as the residual are defined exactly as in Section 4.2. With this notation in hand, we now assess how inequality in the U.S. has evolved over time.

Table 5: Industry-enhanced variance decomposition

	Interval 1: 1996-2002	Interval 2: 2004-2010	Interval 3: 2012-2018	Growth: 1 to 3
Total variance	0.794	0.862	0.915	0.121
Between-firm, within-industry	14.0%	14.7%	15.3%	23.1%
Firm segregation	6.3%	6.6%	7.0%	11.6%
Firm pay premium	3.1%	3.4%	3.1%	2.9%
Firm sorting	4.6%	4.7%	5.1%	8.6%
Between-industry	21.4%	23.6%	26.8%	61.9%
Industry segregation	7.4%	7.8%	9.7%	25.2%
Industry pay premium	4.2%	4.8%	4.8%	8.7%
Industry sorting	9.9%	11.0%	12.3%	28.0%
Within-firm	64.6%	61.7%	58.0%	14.9%
Person effect and observables	48.2%	46.5%	44.3%	18.8%
Residual	16.4%	15.2%	13.7%	-3.9%

Notes: Persons with annual real earnings > \$3770 in EINs with 20 or more employees. See Equation (7) for definitions.

5 Within- and between-industry sorting, segregation, and pay premia

5.1 Estimates of the industry-enhanced variance decomposition

Table 5 exploits our industry-enhanced AKM decomposition to understand rising earnings inequality using person, firm, and covariance effects.[18] The first three columns of Table 5 show results for our three intervals while the last column computes the terms underlying the change in inequality from our first to last intervals (1996-02 to 2012-18).

Between-industry dispersion accounts for 61.9% of the growth in inequality over the time period covered by our dataset. 28.0% of the total rise in inequality can be attributed to rising between-industry sorting. Just over one-fourth (25.2%) of the total rise in inequality can be attributed to increasing between-industry segregation. These two estimates imply that changes in industry-level sorting and segregation account for more than half ($28.0\% + 25.2\% = 53.2\%$) of the total rise in

[18] Appendix Table A1 presents estimates of Equation (6), and Appendix Table A2 aggregates these estimates into firm-level segregation, pay premium, and sorting. Appendix Table A3 presents estimates of Equation (7) before aggregating them as done in Table 5.

Table 6: Sources of between-industry variance growth, top 30 industries

Industry relative earnings	Number of industries	Total contribution to between-industry variance growth	Share of contribution explained by between-industry: segregation	pay premium	sorting
High-paying	19 industries	0.041	42.3%	13.8%	43.9%
Low-paying	11 industries	0.033	36.5%	15.8%	47.7%

Notes: Persons with annual real earnings > \$3770 in EINs with 20 or more employees. See Equation (7) for definitions.

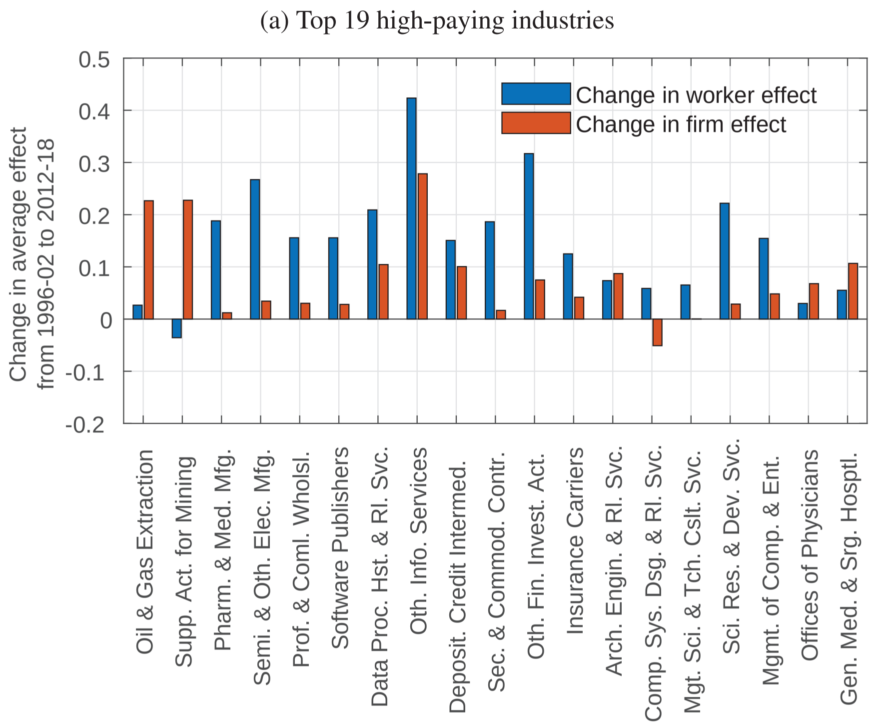
inequality during the time period we consider. Highly paid workers increasingly work in the same industry as other highly paid workers, and in high-paying industries such as Software Publishers (5112).[19] Analogously, workers who command a low wage increasingly work among each other and in low-paying industries such as Restaurants and Other eating Places (7225). Rising dispersion in industry-level pay premia account for a smaller but still substantial 8.7% of the total rise in inequality.

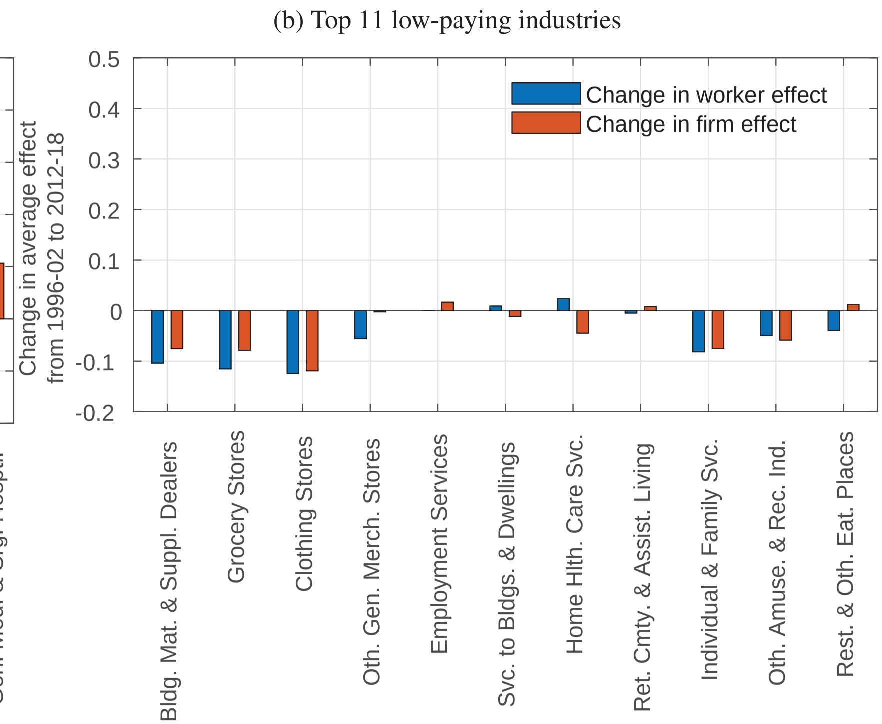
To explore the between-industry contribution to increasing inequality further, Table 6 presents the between-industry sorting, segregation, and firm pay premia contributions for the nineteen high-paying and eleven low-paying industries among the dominant thirty industries. All of these components contribute substantially to rising between-industry dispersion. Segregation is relatively more important in the nineteen high-paying industries (42.3% vs. 36.5%) while sorting is relatively more important in the eleven low-paying industries (47.7% vs. 43.9%). These findings highlight that there are subtle but important differences in the nature of the restructuring of the organization of work at the top and bottom earnings industries. For the top earnings industries, reorganizations have concentrated high person effect workers together in the same industries. For the bottom earnings industries, ongoing changes have led low person effect workers into industries with especially low firm premia.

Some broad patterns on the sources of increasing between-industry inequality due to sorting, segregation, and pay premia emerge from Figure 2. This figure shows the change in the average worker effect (including the effects of observable characteristics, which are normalized by $\bar{y}^p$ in each time period) and firm effect for each of our top thirty industries, each normalized to mean zero. The average firm effect $\bar{\psi}^k$ for each industry k is that industry's pay premium. The sum of the average worker effect $\bar{X}^k\beta + \bar{\theta}^k$ and average firm effect is equal to that industry's average pay, that is, $\bar{y}^{k,p} = \bar{X}^k\beta + \bar{\theta}^k + \bar{\psi}^k$. Recall that we report the average of $\bar{y}^{k,p} - \bar{y}^p$ and its change over time for each of our

[19]The contributions of the industries mentioned in this paragraph to sorting and segregation are presented in Appendix Table A5, which also lists the sorting, segregation, and firm premia contributions for the top ten contributing industries.

Figure 2: Change in industry-level average worker and firm effects





Notes: Persons with annual real earnings > $3770 in EINs with 20 or more employees. Changes in the average worker and firm effect are calculated after normalizing relative to the average worker effect and firm effect, respectively.

top thirty industries in Table 3. The components of Figure 2 for each industry sum to these changes in Table 3.

In 23 of the top 30 industries, the change in the average worker and industry premia have the same sign. The correlation in the top 30 industries between the changing industry premia and changing average worker effects is 0.55. Notable contributors to this correlation are high-paying industries like Other Information Services with a very large increase in the average worker effect of 42.3 log points (52.7%) and the industry premia of 27.8 log points (32.1%) and low-paying industries like Grocery Stores with a large decline in the average worker effect of 11.5 log points (12.2%) and the industry premia of of 7.9 log points (8.2%). These patterns underlie the substantial contribution of sorting.

For an industry's contribution to inequality to have a large segregation component, it must exhibit especially large changes in its average worker effect. In eight of our nineteen high-paying industries (Figure 2(a)), the change in average worker effects contributes more than 10 log points (10.5%) to that industry's change in average earnings, while the industry effect contributes less than 5 log points (5.1%). These include the two manufacturing industries, as well as Professional and Commercial Equipment and Supplies Merchant Wholesalers (4234), and Management of Companies and Enterprises (5511).

What firm-level inequality is left after we account for industry-level differences? Here we return to Table 5. In the cross-section, less than one-sixth (14.0% to 15.3%) of earnings dispersion occurs between firms in the same industry. Looking at growth, we find that 23.1% of variance growth is between firms, within industries. Of this, segregation accounts for 11.6% of the overall increase in inequality, while firm-level sorting accounts for 8.6%. Rising within-industry, between-firm pay premia play a smaller but nontrivial role in rising inequality, and account for 2.9% of the increase in inequality.[20]

Table 5 also describes within-firm inequality. In the cross section, most of the variation in earnings is within-firm rather than between-firm – but notably the share is declining from 64.6% in the first interval to 58.0% in the last. Although its share of overall earnings dispersion falls over time, rising within-firm inequality accounts for a modest amount (14.9%) of the growth in inequality. This mostly reflects an increase in the dispersion of worker effects (18.8%), as the residual has a relatively small role offsetting inequality growth (-3.9%). These estimates are quite close to analogous results reported by Song et al. (2019) for a similar time period. We elaborate on this in appendix section D.

[20]Details about industry contributions to the between-firm, within-industry contributions are provided in Appendix B.

Table 7: Changes in employment and earnings, by industry earnings, mega firms vs. others

Industry relative earnings	Number of industries	Firm employ-ment	Employment share: average	Employment share: change	Relative earnings: average	Relative earnings: change
30 industries with variance contribution $>$ 1%						
		Any	21.1%	2.2%	0.440	0.177
High-paying	19 industries	10,000+	3.8%	1.4%	0.576	0.145
		<10,000	17.3%	0.8%	0.410	0.174
		Any	18.1%	6.0%	-0.586	-0.069
Low-paying	11 industries	10,000+	4.3%	2.5%	-0.492	-0.125
		<10,000	13.8%	3.5%	-0.613	-0.061
271 industries with variance contribution $\leq$ 1%						
		Any	34.9%	-6.8%	0.281	0.046
High-paying	146 industries	10,000+	3.9%	-1.2%	0.646	0.042
		<10,000	31.0%	-5.7%	0.236	0.052
		Any	25.9%	-1.3%	-0.325	-0.002
Low-paying	125 industries	10,000+	3.3%	-0.5%	-0.404	-0.061
		<10,000	22.6%	-0.9%	-0.314	0.006

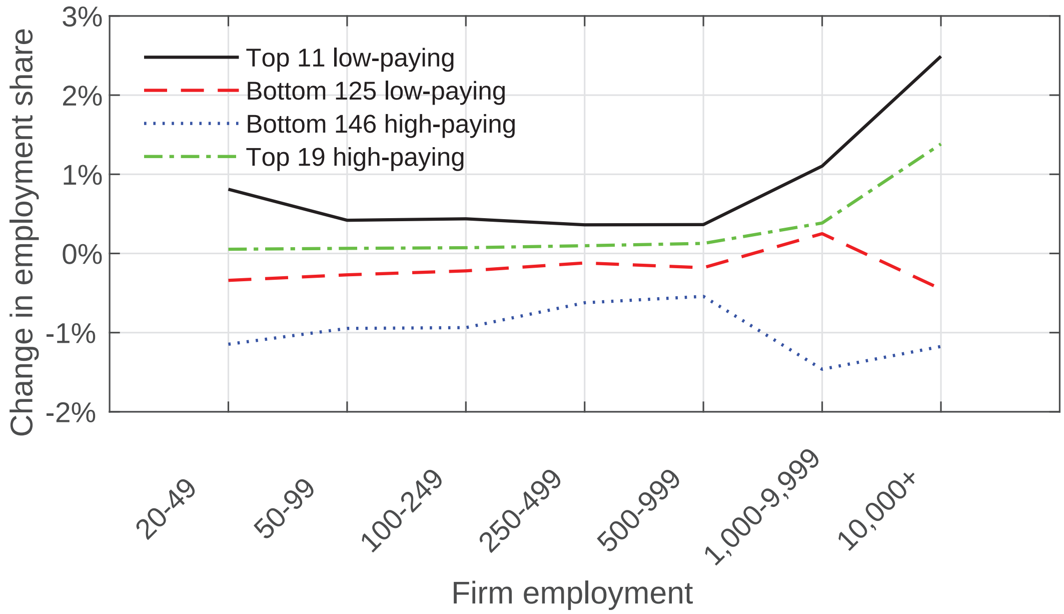
Notes: Persons with annual real earnings > $3770 in EINs with 20 or more employees. Averages and changes use the employment shares and earnings from the 1996-2002 and 2012-2018 intervals. Average log earnings for group relative to the economy average.

6 Mega firms

In this Section, we show that changes in the employment shares and size-earnings premia for mega (10,000+) firms play a critical role in accounting for rising between-industry earnings inequality.[21] Table 7 shows descriptive statistics of employment and earnings in mega firms and non-mega firms in our four industry groups. One immediate result in Table 7 is that employment has shifted over time to the top thirty industries. The employment share of the top thirty industries increased by 8.2 percentage points, with most of this increase (6.0 percentage points) among the eleven low-paying industries. The employment share of the other 271 industries analogously declined by 8.2 percentage points, with most of this decline (6.8 percentage points) among the 146 high-paying industries.

[21] Song et al. (2019) also examine the role of mega firms but with a different focus. They do not explore the close connection between rising between-industry dispersion and mega firms in a relatively narrow set of industries. They do note however that rising within-firm inequality is greater at the mega firms. We find in Appendix B that the industries that contribute most to rising between-firm, within-industry inequality.

Figure 3: Change in employment share by size class, by industry group



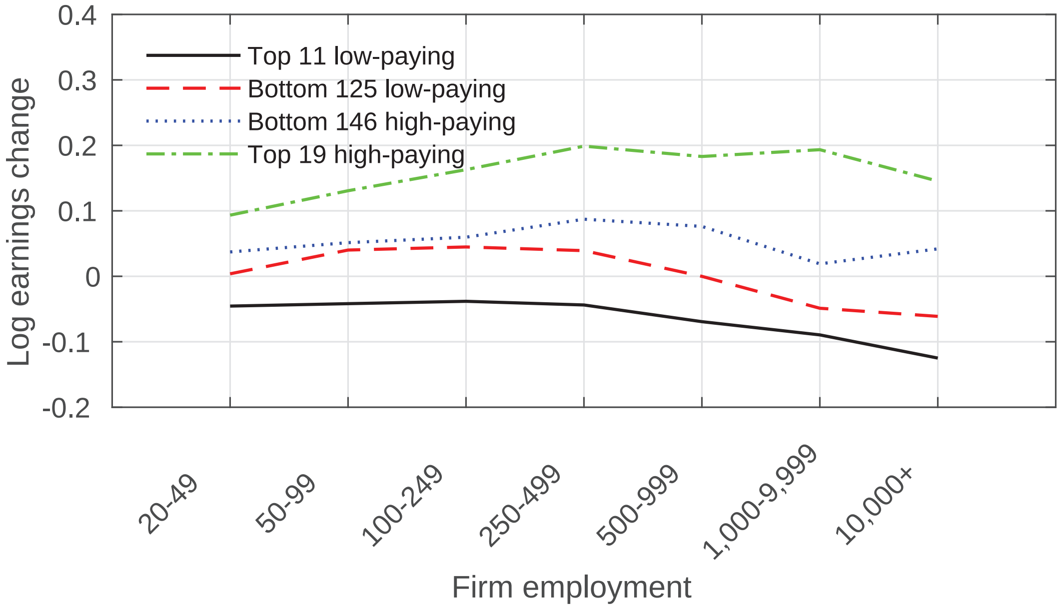
Notes: Persons with annual real earnings > $3770 in EINs with 20 or more employees. Changes in the employment shares are expressed in terms of percentage points. The denominator is total employment across all size classes and industry groups.

The substantial increase in the employment share of the top thirty industries is driven by mega firms.[22] This is evident in both Table 7 and Figure 3. The employment share of the eleven low-paying industries increased in every size class, with mega firms exhibiting the largest increase (2.5 percentage points). The nineteen high-paying industries had a smaller increase in employment, but most of this increase (1.4 percentage points of the 2.2 percentage point total) is accounted for by mega firms. Given the high average relative pay of mega firms in the high-paying industries (57.6 log points, or 77.9%) and the low average relative pay of mega firms (-49.2 log points, or -63.6%) in the low-paying industries, these shifts in employment to mega firms contributed to rising between-industry earnings inequality.

Mega firms also play a key role in the changing earnings patterns of the top thirty industries that contribute to rising between-industry inequality. For the eleven low-paying industries, the relative pay of mega firms decreased by 12.5 log points (13.3%) compared to a decline of 6.1 log points (6.3%) for the non-mega firms. Both mega firms and non-mega firms in the nineteen high-paying

[22]Figure 3 shows the change in employment share by detailed size class for each of our four industry groups. The corresponding employment share levels in the first interval (1996-2002) and in the third interval (2012-2018) are given in Appendix Figure A3.

Figure 4: Change in log earnings by size class, by industry group



Notes: Persons with annual real earnings > \$3770 in EINs with 20 or more employees.

industries exhibit large earnings increases: 14.5 log points (15.6%) for mega firms and 17.4 log points (19.0%) for non-mega firms. Earnings at mega firms increased relative to the smallest firms in the top-paying industries but not by as much as the increase in relative earnings at large but not mega firms.[23] In contrast, relative earnings increases at mega firms in the 146 remaining high-paying industries are modest (4.2 log points, or 4.3%) compared to 14.5 log points in the top nineteen high-paying industries. Similarly, relative earnings declines at the mega firms in the remaining 125 low-paying industries are modest (-6.1 log points, or -6.3%) compared to the -12.5 log points in the top eleven low-paying industries.

Figure 4 shows the change in log earnings by size class and industry group. For the nineteen high-paying industries, the size premium rises for all size classes relative to the smallest size class between our first interval (1996-2002) and our third interval (2012- 2018).[24] For these industries, earnings rise by 19.9 log points (22.0%) for size class 250-499, by 18.3 log points (20.1%) for size class 500-999,

[23] Appendix Figure A5 shows the cross-sectional size-earnings premia for the 1996-2002 and the 2012-2018 intervals. Among the top nineteen high-paying industries, the size-earnings profile shifts upward, with increases in all size classes.

[24] Appendix Figure A6 shows the cross-sectional size-earnings premia of the AKM components for our dominant thirty industries. Appendix Figure A7 shows the decomposition of the changing size-earnings premium into its AKM components. Appendix Figure A8 shows the earnings changes and employment share changes for all industries pooled together. We find an inverted U-shaped change in the size-earnings premium, with particularly large declines for mega firms. This is accompanied by a rising share of employment at mega firms (2.2 percentage points). The pooled results are consistent with Bloom et al. (2018) but pooling across all industries masks the disproportionate role of the top 30 industries and the different patterns for the high-paying industries among the top 30.

by 19.3 log points (21.3%) for size class 1000-9999, and by 14.5 log points (15.6%) for mega firms. The eleven low-paying industries exhibit a decline in the size-earnings premium over time, with a steeply declining size premium for the larger firms and the mega firms (Figure A7). The decline at mega firms in the eleven industries is 12.5 log points (13.3%). This decline represents a flattening of the size-earnings premium. This decline in the size-earnings premium is accompanied by a sharp increase in employment in mega firms.

7 Industries and occupations

We now explore the role of occupation in the between-industry differences using the Occupational Employment and Wage Survey (OEWS). Using published aggregates of OEWS data, we obtain a consistent time series for 22 occupation categories for 281 detailed (4-digit) NAICS industries for the years 2002-2016.[25]

Our OEWS dataset allows us to compare our top 30 industries with 254 others, including 141 of the 146 other high-paying industries and 110 of the 125 other low-paying industries explored above. We have the employment share and average earnings for each industry by occupation group. We first use this to explore the growth in between-industry earnings inequality. Analogous to our findings above, the top 30 industries account for 96.2% of rising between-industry inequality growth. Our top 19 high-paying industries account for 64.5% of between-industry growth, and our top 11 low-paying industries account for 31.7%. The other 141 high-paying industries and 110 low-paying industries account for only 2.5% and 1.3%, respectively.

Using a similar methodology as used above with the AKM decomposition (see appendix E), the contribution of the between-industry effects can be decomposed into the contribution of industry premia, sorting of high (low) wage occupations into high (low) wage industries, and segregation of high (low) occupations across industries (see Table 8). For both the 19 high paying and 11 low paying industries, the between-industry contribution is dominated by sorting effects accounting for 52.0% and 50.1%, respectively. Segregation and industry premia both make a substantial contribution. Compared to the analogous results using the AKM decomposition in Table 8, segregation plays a smaller role and industry premia a larger role. Appropriate caution is needed here as this is an apples-to-oranges comparison (LEHD micro data vs OEWS industry by occupation data), but it does suggest

[25]For details about our dataset construction, see Appendix E.

Table 8: Sources of between-industry variance growth, top 30 industries, using OEWS

Industry relative earnings	Number of industries	Share of contribution to between-industry variance growth	Share of contribution explained by between-industry: segregation	pay premium	sorting
High-paying	19 industries	64.5%	25.3%	22.7%	52.0%
Low-paying	11 industries	31.7%	31.7%	18.2%	50.1%

Notes: Authors' calculations of published OEWS aggregates.

an important role for sorting and segregation of occupations in accounting for the dominant role of between-industry inequality.[26]

We interpret the especially large role of sorting as highlighting the importance of both industry and occupation. This sorting component only arises in the presence of industry effects that are not accounted for by occupation. The between-industry segregation component also highlights the importance of the interaction of industries and occupations via the increased segregation of specific occupations in specific industries.

To get a sense of the contribution of occupations to industry-level wage inequality, we divide occupations into two broad categories, 12 that are high-paying relative to the overall average, and ten that are analogously low-paying.[27] The results are shown in Table 9. Occupational shifts appear to occur, at least broadly, at the industry level. Nearly all of the growth in the employment share of high-paying occupations occurs in the 19 high-paying industries, and all of the growth in low-paying occupations occurs in our 11 low-paying industries. Our 19 high-paying industries increase their employment share by 1.3% in the OEWS data. This reflects strong growth among our 12 high-paying occupations, whose share of employment grows by 2.0%. The employment share of our 10 low-paying occupations in these 19 high-paying industries contracts by 0.7%. The 11 low-paying industries increase their employment share by 3.1%. This mostly reflects an increase among our 10 low-paying occupations, whose share of overall employment increases by 2.8 percentage points.

[26] We focus on the top 30 industries in Table 8. For all industries, we find that industry segregation accounts for 36% of the rising between-industry inequality, industry premia 3% and industry sorting 61%.

[27] Our high-paying occupations are, by two-digit Standard Occupation Classification group: Management (11), Business and Financial Operations (13), Computer and Mathematical Science (15), Architecture and Engineering (17), Life, Physical, and Social Science (19), Community and Social Services (21), Legal (23), Education, Training, and Library (25), Arts, Design, Entertainment, Sports, and Media (27), Healthcare Practitioner and Technical (29), Construction and Extraction (47), and Installation, Maintenance, and Repair (49). Our low-paying occupations are Healthcare Support (31), Protective Service (33), Food Preparation and Serving Related (35), Building and Grounds Cleaning and Maintenance (37), Personal Care and Service (39), Sales and Related (41), Office and Administrative Support (43), Farming, Fishing, and Forestry (45), Production (51), and Transportation and Material Moving (53).

Table 9: Changes in employment and earnings, by industry earnings, by occupation groups

Industry group earnings	Occupation group earnings	Employment share: average	Employment share: change	Relative earnings: average	Relative earnings: change
30 industries with variance contribution $>$ 1%					
	All	16.3%	1.3%	0.405	0.052
19 high-paying	12 high-paying	9.7%	2.0%	0.707	0.067
	10 low-paying	6.6%	-0.7%	-0.039	0.029
	All	19.8%	3.1%	-0.437	-0.009
11 low-paying	12 high-paying	2.1%	0.3%	0.269	0.031
	10 low-paying	17.7%	2.8%	-0.521	-0.013
251 industries with variance contribution $\leq$ 1%					
	All	30.0%	-2.8%	0.213	-0.018
141 high-paying	12 high-paying	13.0%	-0.1%	0.496	0.005
	10 low-paying	17.1%	-2.7%	-0.002	-0.036
	All	33.9%	-1.6%	-0.130	-0.019
110 low-paying	12 high-paying	12.8%	-0.2%	0.238	-0.017
	10 low-paying	21.1%	-1.5%	-0.353	-0.020

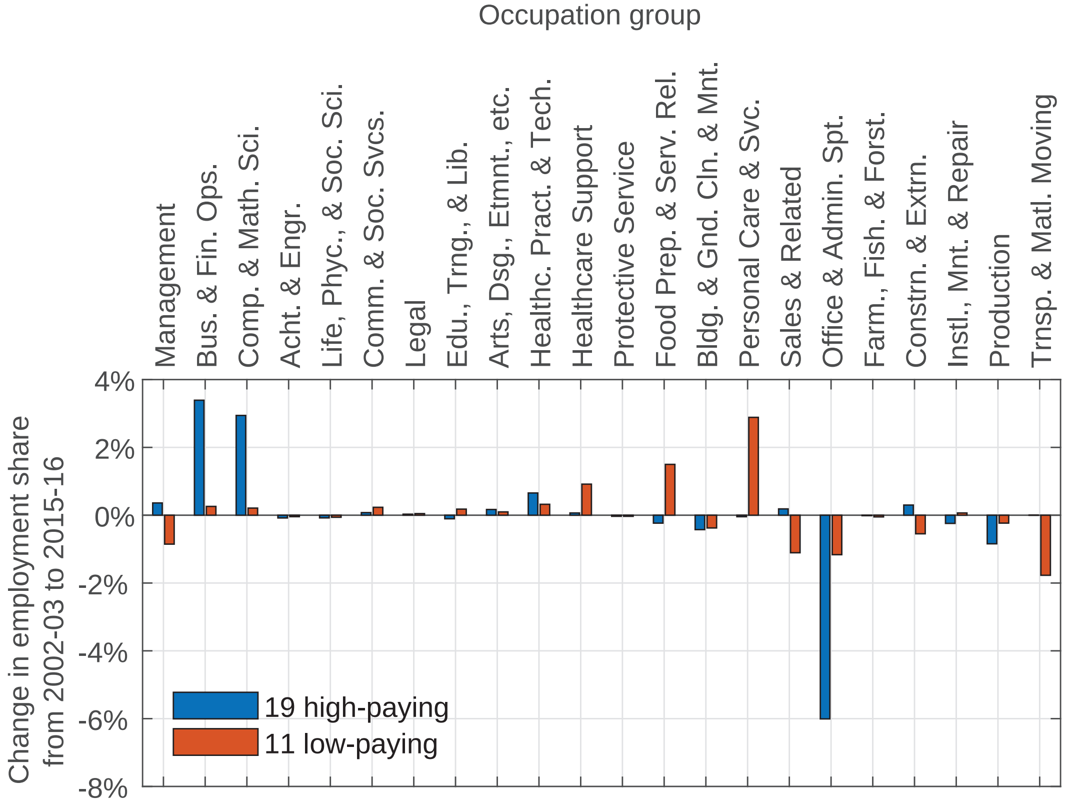
Notes: Authors' calculations from OEWS published aggregates. Changes in employment and earnings compare year intervals 2002-2003 with 2015-2016. Industry pay designations and contributions to variance growth follow the LEHD administrative records data.

Outside the 30 industries that contribute to inequality, others have declining employment in low-paying occupations. The employment share of our other 141 high-paying industries declines by 2.8 points, and nearly all of this (2.7 percentage points) occurs among low-paying occupations. The employment share of the 110 low-paying industries likewise decline by 1.6 percentage points, which almost matches (1.5 percentage points) the decline in the employment share of low-paying occupations.

Table 9 also illustrates the role of industry-occupation pay differentials in rising inequality. Our 19 high-paying industries have a strong (40.5 log point) earnings differential. This reflects an even larger (70.7 log point) differential among our 12 high-paying occupations in these industries. These industries also had the highest earnings growth, both overall (5.2 log points), and especially among high-paying occupations (6.7 log points). Workers in our 10 low-paying occupations in these 19 high-paying industries had earnings gain as well (2.9 log points).

Earnings changed by relatively less among our top 11 low-paying industries. The overall change

Figure 5: Change in employment by occupation group



Notes: Authors' calculations of published OEWS aggregates.

in earnings was a decline of only 0.9 log points. This reflects a gain in earnings of 3.1 log points among the relatively rate high-paying occupations. Low-paying occupations in these industries had an earnings decline of 1.3 log points. There were somewhat larger earnings declines among the other 251 industries (1.8 log points among high-paying industries, 1.9 log points among low-paying industries), and these largely reflect declines in the earnings of low-paying occupations (of 3.6 log points, and 2.0 log points, respectively).

We drill down further into our occupation data in Figure 5. Here, we explore changes in occupation at a finer level, considering employment changes across all 22 occupation groups. We focus on our top 30 industries as these are where the growth in high-paying and low-paying occupations are concentrated. There are substantial differences in the changing mix of occupations across the top 19 and bottom 11. The top 19 industries have large increases in business and financial operations and computer and math oriented occupations. Both are very high wage occupations. In the top 19 industries, there are large declines in office and administrative support, a low wage occupation. Other nontrivial changes include increases in management (high wage) and decreases in production (low wage). The employment share of the bottom 11 industries increases, with particularly strong growth in food preparation and personal care and services (very low wage occupations). These low-paying industries also exhibit a nontrivial decline in management occupations.

One potential issue is whether greater occupational detail may change the inferences in this section. However, this concern neglects the high degree of concentration of occupations across detailed industries. In appendix figure E3, we show that even for the 2-digit occupations we use in the current analysis, occupations are highly concentrated in detailed industries. In this figure, we depict the share of employment in each occupation accounted for by the top 20 industries employing that occupation. For the median occupation, this share exceeds 80 percent. For a few occupations such as management, production workers, and transportation workers, this share is as low as 40 percent. Using even more detailed occupations yields even higher concentration. Appendix figure E4 shows that using 3-digit occupation, the median occupation top 20-industry concentration ratio of employment exceeds 90 percent and using 6-digit occupation, the median occupation has a top 20-industry concentration ratio of 100 percent. This degree of concentration of occupations by industry is consistent with our narrative that industries help differentiate how production is organized differently across firms. Detailed occupations which are associated with specific detailed industries provide guidance about how the skills and tasks within these industries are accomplished.[28]

8 Industry premia, sorting and segregation: robustness

Our analysis of the OEWS provides guidance about the interaction of the role of occupation and industry for rising inequality. It also serves as a form of robustness analysis for our characterization of the role of industry premia, sorting and segregation. The AKM based analysis is more comprehensive using person-level data but the OEWS findings highlight that a dominant role of between-industry sorting and segregation is robust to this alternative source of data and methodology.

Using linked CPS-LEHD data (developed and discussed in detail in our companion paper, Haltiwanger, Hyatt and Spletzer (2022)), we provide in this section a human capital approach as an alterna-

[28]Using more detailed occupation can influence the decomposition of the between industry components into sorting, segregation and industry premia. The public domain OEWS data does not readily support such analysis given suppressions and concordance issues of occupation code changes in the public domain OEWS data. Our concentration analysis in this section using more detailed occupations in Appendix Figure E4 uses one year of data that mitigates these issues (since suppressions and concordance issues interact).

Table 10: Variance decomposition using an earnings equation: AKM vs. human capital

Data Source Specification	LEHD AKM	CPS-LEHD linked AKM	CPS-LEHD linked human capital
Between-industry:	61.9%	66.2%	66.2%
Segregation	25.2%	26.9%	14.8%
Pay premia	8.7%	9.6%	22.0%
Sorting	28.0%	29.6%	29.4%

Notes: The first column is taken from Table 5. The second and third columns are from our companion paper using the linked CPS-LEHD dataset, common-coded sample. 299 4-digit NAICS industries. The three components add up to the total between-industry contribution. There are some small differences in the sample restrictions for the CPS-LEHD dataset to make it more comparable to the Hoffman, Lee, and Lemieux (2020) approach relative to the full LEHD datasets used in this paper (that targeted restrictions to be comparable to Song et al. (2019)). However, the results in columns 1 and 2 show these differences are not important quantitatively for the decomposition of rising between-industry inequality.

tive to AKM to quantify the role of industry premia, industry segregation and industry sorting. Specifically, using the linked CPS-LEHD data, we estimate a standard human capital equation building on Hoffman, Lee, and Lemieux (2020) (hereafter HLL) relating earnings to observable characteristics including age, education, occupation and industry but critically use detailed LEHD industry codes instead of CPS codes. For current purposes, we also use LEHD earnings to make results comparable to those above. [29] With these estimates, we decompose the increase in earnings inequality into within industry (the contribution of increasing dispersion in earnings within industries across workers based on human capital observables as well as the residual), segregation (the contribution of human capital observables to changing dispersion across industries), sorting (the increasing covariance of the contribution of human capital observables and industry premia), and industry premia (see appendix G).

Table 10 compares the decomposition of the contribution of industry using the full LEHD with AKM as in Table 5 with the linked CPS-LEHD analysis using the human capital equation approach. In taking advantage of the data infrastructure from both approaches, we can also integrate the AKM estimates from the full LEHD analysis into the linked CPS-LEHD.

In comparing columns 1 and 2 in Table 10 we find that the overall contribution of industry and in turn the respective contributions of industry premia, sorting and segregation using AKM estimates

[29] In the companion paper, we report results using CPS earnings and show that most of increase in CPS earnings dispersion is due to between industry effects once one uses LEHD industry codes and takes into account sorting and segregation contributions.

are very similar in the full LEHD and the linked CPS-LEHD. This finding is reassuring as it implies that the CPS-LEHD linked data can replicate the core patterns from the LEHD.

In comparing columns 2 and 3 in Table 10 we find that whether using AKM based estimates or human capital equation based estimates that the contributions of industry premia, segregation and sorting are broadly similar. By construction, the overall contribution of industry is the same in columns 2 and 3.[30] Interestingly, this overall contribution is similar to that using the full LEHD. The sorting component of industry is almost identical whether using AKM or human capital estimates. Industry premia appears to be especially important when we use the human capital approach. In contrast, when we use AKM, between-industry segregation becomes more important.

9 Industry premia, sorting and segregation: discussion

We find that between-industry sorting is the most important component in accounting for the rising between-industry inequality whether we use AKM, a human capital equation approach or jointly estimated industry and occupation effects from the OEWS data. The positive between-industry sorting contribution to inequality arises because of three factors. First, there are between-industry premia – in their absence there is no contribution of sorting as defined in this paper. Second, these premia are positively related to earnings differences across industries from worker characteristics such as age, education, occupation or person effects. Third, this positive covariance is increasing over time. Our estimates imply that increasing between-industry sorting accounts for at least 25% of the overall increase in inequality. The combined effect of increased sorting (that depends on the presence of industry premia) with the direct effect of rising dispersion in industry premia accounts for at least 35% of the overall increase in inequality.

Our findings imply that whatever the mechanism driving increasing inequality it must have a prominent role for industry as well as a prominent role for industry premia and between-industry sorting interacting with the mechanism. Conceptually, industry premia (and more generally firm premia) inherently arise due to frictions or market imperfections (e.g., monopsony power) in labor markets. While we acknowledge there are issues for estimating these conceptual effects, the robustness of our findings across different source data and methodology mitigate such concerns.

[30] Columns 2 and 3 use LEHD earnings in the CPS-LEHD linked data. When using CPS earnings we find that the between-industry contribution is similar: 65.5%.

Viewed from this perspective, our findings rule out explanations of rising inequality that don't have a prominent role for industry premia through both direct and sorting effects. To help appreciate this, consider a recent pathbreaking paper by Acemoglu and Restrepo (2022) that documents that industry plays a prominent role in accounting for rising inequality. Their findings highlight differential adoption of advanced technologies across industries along with technology-skill complementarities. The structural model they use to interpret these empirical findings emphasizes that technology adoption displaces some tasks conducted by less skilled labor but is complementary with tasks conducted by more skilled labor. Their finding of an important role for industry is broadly consistent with our finding of the dominant overall role of rising between-industry inequality. However, the structural model has competitive labor markets without frictions so that their model implies that the prominent role of industry they find is associated with rising segregation between industries without any role for industry premia and sorting. In other words, their model focuses on the changing composition of workers across industries without any interaction of that changing composition with industry premia (i.e., what we have defined as sorting in this paper). Our empirical results suggest that enhancing the structural model to permit frictions and imperfect competition in labor markets is critical for capturing the variation in the data. This argument is not restricted to this specific mechanism (which is a form of skill biased technical change). Our results imply that any mechanism must account for the combined contribution of increased between-industry sorting and rising between-industry premia.

The contribution of between-industry segregation is also important and informative. First, this contribution highlights the mechanism must have an industry component of changing contribution of workers across industries. Second, there are interesting open questions about the consequences of increasing segregation. Workers learn from co-workers so that segregation may limit such learning with especially adverse effects for lower skilled workers. Relatedly, economic mobility of lower skilled workers may be adversely impacted.

10 Conclusion

Rising earnings inequality is dominated by rising between-firm inequality. Our analysis as well as the recent literature emphasizes that this largely reflects how firms are organizing themselves in terms of their workforce. High (low) earnings workers are more likely to work with each other (increased segregation), and high (low) earnings workers are more likely to work at high (low) premia firms (sorting).

Our contribution is to highlight the dominant role of industry effects in accounting for this structural change of how firms organize their workforces. Most of rising between-firm inequality is accounted for by rising between-industry dispersion in earnings. The between-industry component accounts for 61.9% of total increasing earnings inequality, and 72.8% of between-firm inequality growth. This changes the narrative of the sorting and segregation contributions. High (low) earnings workers are more likely to work with each other in specific industries and high (low) earnings workers are more likely to work in high (low) average firm premia industries. This inference is robust to identifying industry premia, sorting and segregation from decomposing earnings using the AKM approach, the standard human capital approach using worker observables including age, education, and occupation, and using high quality industry by occupation data.

Not only do industry effects dominate but it is a relatively small share of industries that account for virtually all the increasing dispersion in earnings across industries. We find that about ten percent of the 301 detailed 4-digit NAICS industries account for almost all of rising between-industry dispersion, while accounting for less than 40% of employment. The ten percent of industries that account for virtually all of the increase are drawn from the top and bottom of the earnings distribution in terms of industry-level averages. For those industries at the top of the earnings distribution, their contribution is dominated by rising inter-industry earnings differentials. For industries at the bottom of the earnings distribution, their contribution is dominated by shifts in employment to these very low-earnings industries. For both sets of industries at the top and the bottom of the earnings distribution, increased sorting and segregation between industries dominates but increased dispersion in between-industry firm premia also plays an important supporting role. Increased sorting is relatively more important for the rising between-industry dispersion from the industries at the bottom of the earnings distribution. In contrast, increased segregation is relatively more important for the rising between-industry dispersion from the industries at the top of the earnings distribution.

The top ten percent of industries that account for virtually all of rising between-industry inequality are not randomly spread across the distribution of industries but concentrated in specific industry clusters in the tails of the earnings distribution. At the high end, dominant industries are drawn from high-tech and STEM intensive industries, finance, mining, and selected industries in health. At the low end, dominant industries are drawn from selected industries in retail and health. Notably absent are the vast majority of industries in manufacturing. The top thirty industries are in industry clusters that have exhibited structural transformations that have been the subject of independent study.

The dominance of industry effects is closely linked to the rising importance of mega (10,000+) firms in the U.S. economy. The increasing share of employment accounted for by mega firms is concentrated in the thirty 4-digit industries that account for virtually all of rising between-industry dispersion. This rising share of employment at mega firms is accompanied by a declining size-earnings premium in the eleven low-paying industries. For mega firms in the nineteen high-paying industries in the top 30, earnings premia rise sharply relative to other industries (albeit not as rapidly as other large but not mega firms in these industries).

We find there is a close connection between changes in the occupation distribution across industries and our top 30 industries. Most of the increase in employment in top-paying occupations is accounted for by our 19 top-paying industries, while all of the increase in employment in low-paying occupations is in our 11 low-paying industries. For the top-paying industries, there is substantial increase in earnings differentials for the top-paying occupations. Likewise there is a substantial decline in earnings differentials for the low-paying occupations in our low-paying industries. About half of the increase in between-industry dispersion is accounted for by increased sorting of high (low) paying occupations into high (low) paying industries. About one third is due to increased segregation of high and low paying occupations across the top 30 industries. The remainder is due to increased industry pay premia. In the top-paying industries, there are especially large increases in the growth of employment in computer and math and business and finance occupations (both high-paying occupations). In the low-paying industries, there are especially large increases in the employment growth of food preparation and personal care and services (very low wage occupations). There is a tight relationship between rising between-industry dispersion in a relatively narrow fraction of industries and the changing structure of occupations in these industries. This relationship deserves our attention as it is key for understanding the rise in overall earnings inequality.

Our findings imply that understanding rising earnings inequality during the last several decades requires understanding the restructuring of how firms organize themselves in a relatively small set of industries. Moreover, since it is the between-industry contribution that dominates, it is the common effects of re-organization across firms in the same industry that matter. Many mechanisms such as changing technology, market structure, and globalization likely underlie rising earnings inequality. The focus of future research on the impact of such changes on rising earnings inequality should be on the uneven and concentrated impact of such mechanisms across industries. In addition, since between-industry sorting and industry premia play important roles, the mechanisms must interact with the factors that yield such effects such as frictions and imperfect competition in labor markets.

Our findings, furthermore, identify the nature of the structural changes in terms of industry, occupations, and the rise of mega firms that underlie rising inequality over the period from the 1996-02 to the 2012-18 period. We don't identify the ultimate driving forces of these structural changes but the concentrated nature of the changes in a relatively narrow set of industries provides considerable scope for such investigation. Autor et al. (2020) document related but distinct structural changes in the rise of superstar firms. They don't provide causal evidence of the driving forces but highlight that globalization and the development of technologies yield a "winner take most" mechanism. We think similar driving forces may be at work underlying our findings. A high priority for future research should be to provide causal evidence for these driving forces with a focus on their role in the top 30 industries driving rising earnings inequality.

References

[1] Abowd, John, Bryce Stephens, Lars Vilhuber, Fredrik Andersson, Kevin McKinney, Marc Roemer, and Simon Woodcock. 2009. "The LEHD Infrastructure Files and the Creation of the Quarterly Workforce Indicators." In *Producer Dynamics: New Evidence from Micro Data*, ed. Timothy Dunne, J.Bradford Jensen and Mark J. Roberts, 149-230. Chicago: University of Chicago Press.

[2] Abowd, John, Francis Kramarz, and David Margolis. 1999. "High Wage Workers and High Wage Firms." *Econometrica* 67(2): 251-333.

[3] Abowd, John M., Kevin L. McKinney, and Nellie L. Zhao. 2018. "Earnings Inequality and Mobility Trends in the United States: Nationally Representative Estimates from Longitudinally Linked Employer-Employee Data." *Journal of Labor Economics*, 36(S1): S183-S300.

[4] Acemoglu, Daron and Pascual Restrepo. 2022. "Tasks, Automation, and the Rise in US Wage Inequality" *Econometrica* (forthcoming).

[5] Autor David, David Dorn, Lawrence F. Katz, Christina C. Patterson, and John Van Reenen. 2020. "The Fall of the Labor Share and the Rise of Superstar Firms." *Quarterly Journal of Economics* 135(2): 645-709.

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

[7] Bloom, Nicholas, Fatih Guvenen, Benjamin S. Smith, Jae Song, and Till von Wachter. 2018. "The Disappearing Large-Firm Premium." *AEA Papers and Proceedings* 108: pp. 317-322.

[8] Bonhomme, Stéphane, Kerstin Holzheu, Thibaut Lamadon, Elena Manresa, Magne Mogstad and Bradley Setzler. 2022. "How Much Should We Trust Estimates of Firm Effects and Worker Sorting?" *Journal of Labor Economics* (forthcoming).

[9] Briskar, Juraj, José Rodríguez Mora, Edoardo di Porto, and Cristina Tealdi. 2020. "It's the Sectors, not the Firms: Accounting for Earnings and Wage Inequality Trends in Italy 1985-2018." Unpublished draft, University of Edinburgh.

[10] Briskar, Juraj, José Rodríguez Mora, Edoardo di Porto, and Cristina Tealdi. 2022. "Decomposition of Italian Inequality." Unpublished draft, University of Edinburgh.

[11] Card, David, Ana Cardoso, and Patrick Kline. 2016. "Bargaining, Sorting, and the Gender Wage Gap: Quantifying the Effect of Firms on the Relative Pay of Women." *Quarterly Journal of Economics* 131(2): 633-686.

[12] Card, David, Jorg Heining, and Patrick Kline. 2013. "Workplace Heterogeneity and the Rise of West German Wage Inequality." *Quarterly Journal of Economics* 128(3): 967-1015.

[13] Card, David, Jesse Rothstein, and Moises Yi. 2022. "Industry Wage Differentials: A Firm-Based Approach." Unpublished draft, University of California, Berkeley.

[14] Cooper, Zack, Stuart Craig, Martin Gaynor, and John Van Reenen. 2019. "The Price Ain't Right? Hospital Prices and Health Spending on the Privately Insured." *Quarterly Journal of Economics*, 134(1): 51-107.

[15] Davis, Steve J., and John Haltiwanger. 1991. “Wage dispersion between and within U.S. manufacturing plants, 1963–1986.” *Brookings Papers on Economic Activity*.

[16] Decker, Ryan A., Aaron Flaaen, and Maria D. Tito. 2016. “Unraveling the Oil Conundrum: Productivity Improvements and Cost Declines in the U.S. Shale Oil Industry,” FEDS Notes March 2016.

[17] Dey, Matthew and Houseman, Susan, and Polivka, Anne. 2006. “Manufascturers’ Outsourcing to Employment Services” SSRN: https://ssrn.com/abstract=959428.

[18] Dunne, Timothy, Lucia Foster, John Haltiwanger, and Ken Troske. 2004. “Wage and Productivity Dispersion in United States Manufacturing: The Role of Computer Investment.” *Journal of Labor Economics* 22(2): 397-429.

[19] Fernald, John. 2014. “Productivity and Potential Output before, during, and after the Great Recession.” *NBER Macroeconomics Annual* 29(1): 1-51.

[20] Foster, Lucia, John Haltiwanger and C. J. Krizan. 2006. “Market Selection, Reallocation, and Restructuring in the U.S. Retail Trade Sector in the 1990s.” *Review of Economics and Statistics* 88(4): 748-758.

[21] Foster, Lucia, John Haltiwanger, Shawn Klimek, C.J. Krizan. 2016. “The Evolution of National Retail Chains: How We Got Here.” In *Handbook on the Economics of Retailing and Distribution*, ed. Emek Basker, 7-37. Northampton, MA: Edward Elgar Publishing.

[22] Fulton, Brent. 2017. “Health care market concentration trends in the United States: Evidence and policy responses.” *Health Affairs*, 36(9): 1530–1538.

[23] Goldschlag, Nathan and Miranda, Javier. 2016. “Business Dynamics Statistics of High Tech Industries.” U.S. Census Bureau Center for Economic Studies Paper No. CES-WP-16-55,

[24] Guimarães, Paulo, and Pedro Portugal. 2010. “A Simple Feasible Procedure to Fit Models with High-Dimensional Fixed Effects.” *The Stata Journal* 10(4): 628-649.

[25] Haltiwanger, John, Henry Hyatt, and James R. Spletzer. 2022. “The Role of Industry in Increasing Earnings Inequality: Reconciling Results from the CPS and the LEHD.” Manuscript in progress, U.S. Census Bureau.

[26] Haltiwanger, John and James Spletzer. 2020a. "Between-firm changes in earnings inequality: the dominant role of industry effects." NBER Working Paper No. 26786, February.

[27] Haltiwanger, John and James Spletzer. 2020b. "Rising between-firm inequality and declining labor market fluidity: evidence of a changing job ladder." Paper presented at NBER/CRIW conference on Measuring and Understanding the Distribution and Intra/Inter-Generational Mobility of Income and Wealth, March.

[28] Hecker, Daniel. 2005. "High-technology employment: a NAICS-based update." *Monthly Labor Review*, 128(7), July 2005.

[29] Hoffman, Florian, David S. Lee, and Thomas Lemieux. 2020. "Growing Income Inequality in the United States and Other Advanced Economies." *Journal of Economic Perspectives* 34(4): 52-78.

[30] Kroszner, Randall S. and Strahan, Philip E. 2014. "Regulation and Deregulation of the US Banking Industry: Causes, Consequences, and Implications for the Future." *Economic Regulation and Its Reform: What Have We Learned?*, edited by Nancy L. Rose, Chicago: University of Chicago Press, pp. 485-544.

[31] Oliner, Stephen, Daniel Sichel and Kevin Stiroh. 2007. "Explaining a Productive Decade." *Brookings Papers on Economic Activity*.

[32] Song, Jae, David J. Price, Fatih Guvenen, Nicholas Bloom, and Till von Wachter. 2019. "Firming Up Inequality." *Quarterly Journal of Economics* 134(1): 1–50.

[33] Walker, Ed. 1997. "The Census Bureau's Business Register: Basic Features and Quality Issues." Paper presented at the Joint Statistical Meetings of the American Statistical Association, Anaheim, CA.

Appendices

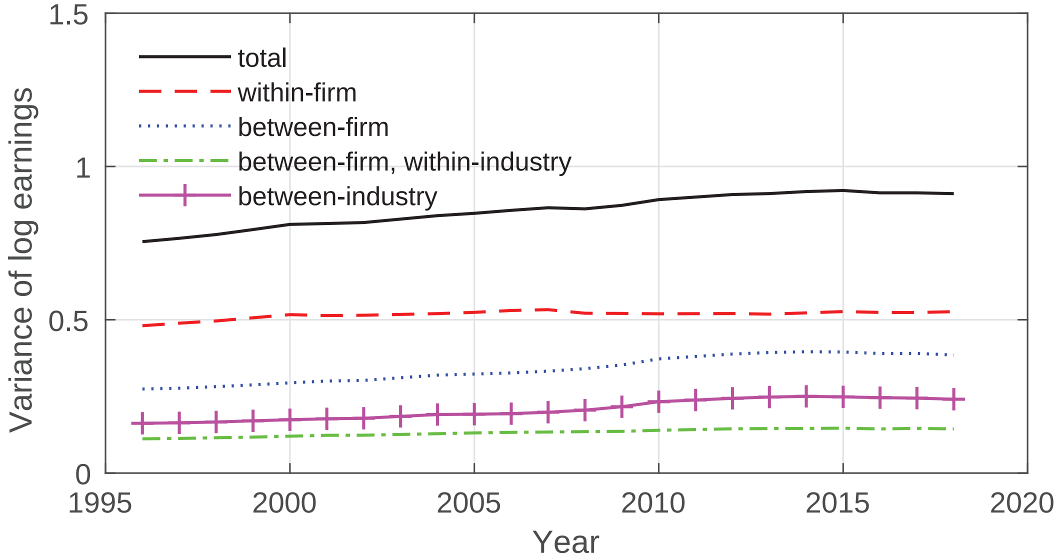
A Supplemental results: pooled males and females

This appendix includes supplemental tables and figures for the results highlighted in the main text (Figures A1 to A8 and Tables A1 to A5).

Figure A1: Descriptive statistics

Notes: Persons with annual real earnings > $3770 in EINs with 20 or more employees. 2003 and 2011 are linearly interpolated.

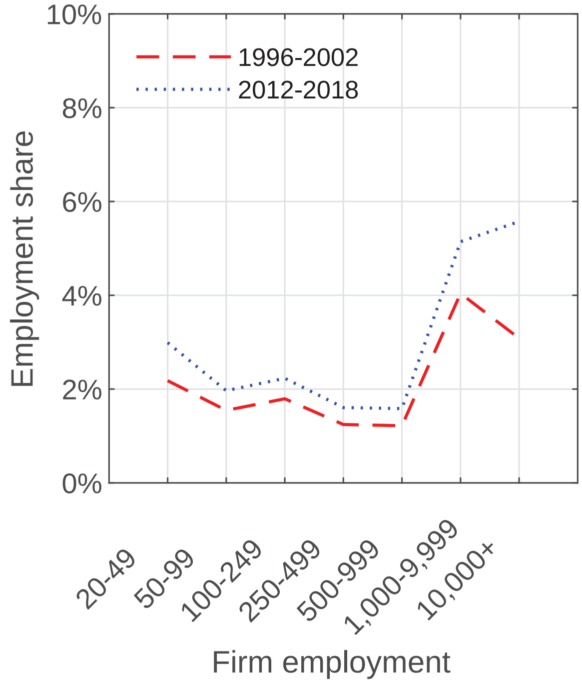
Figure A2: Variance decomposition by year, 1996-2018



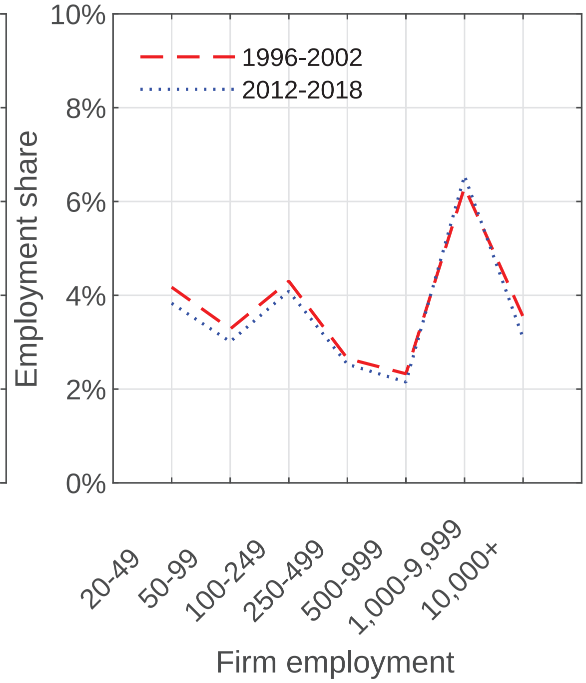
Notes: Persons with annual real earnings > \$3770 in EINs with 20 or more employees. 2003 and 2011 are linearly interpolated.

Figure A3: Employment share by size class and industry group

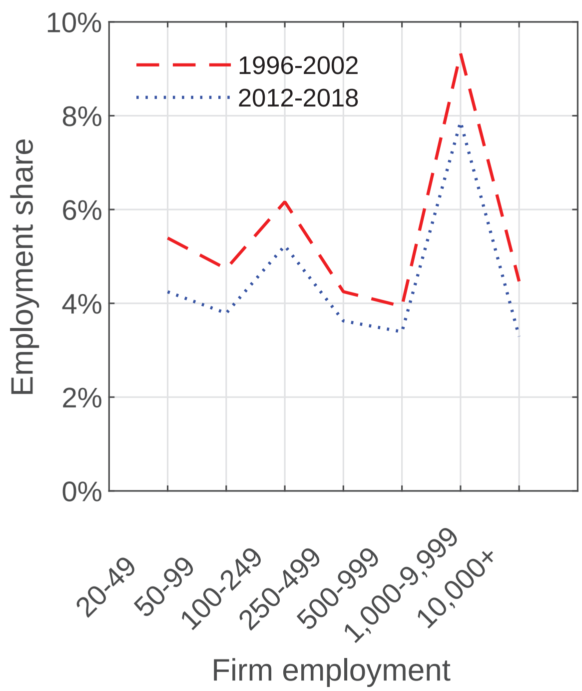
(a) Top 11 low-paying industries



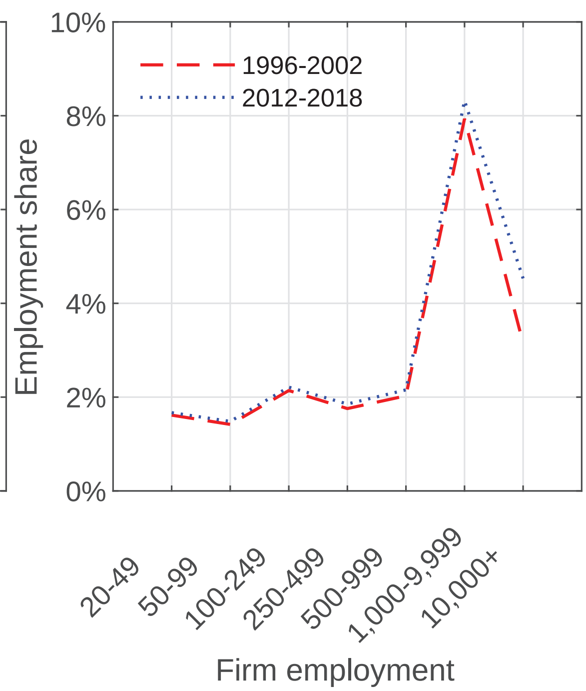
(b) Bottom 125 low-paying industries



(c) Bottom 146 high-paying industries

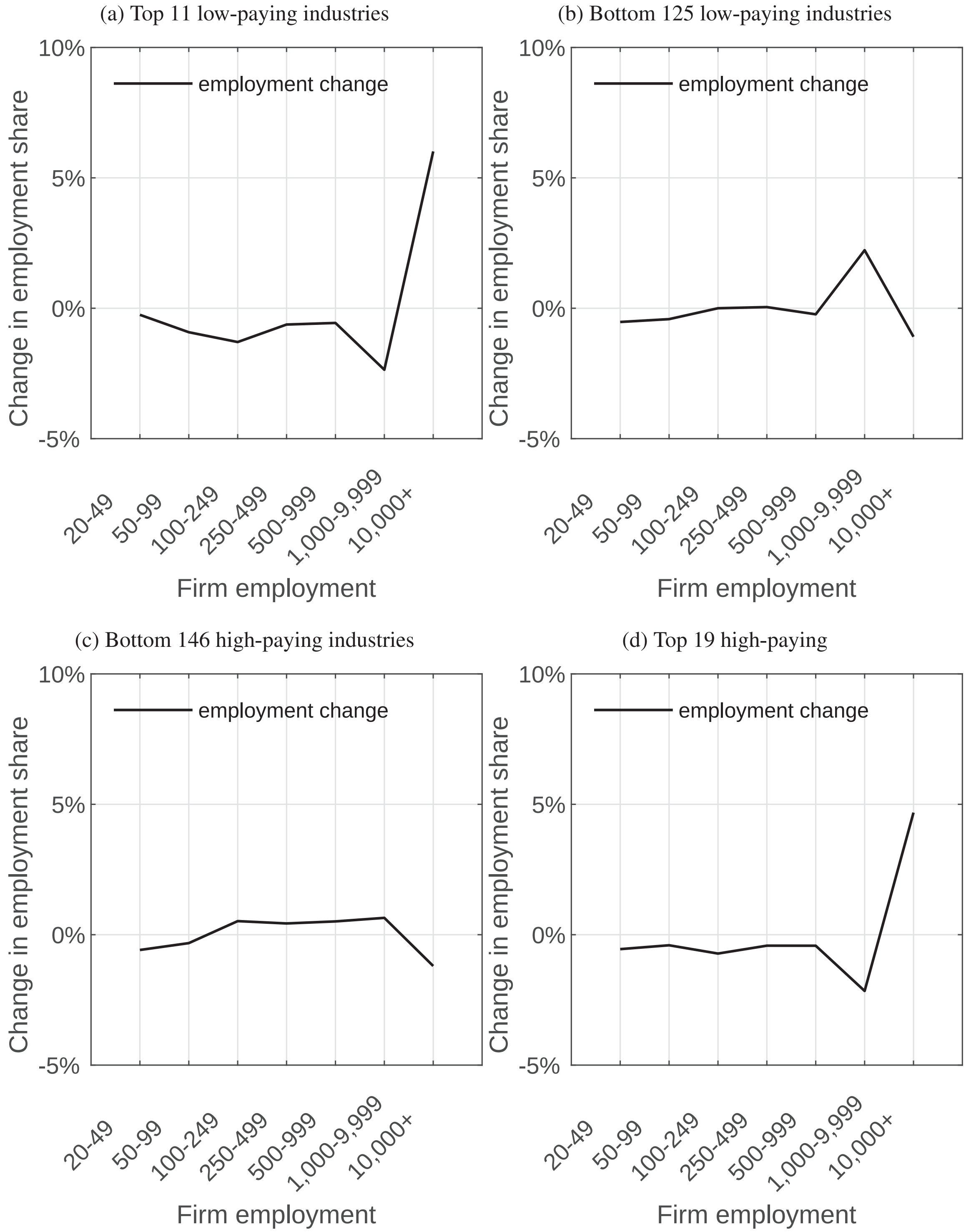


(d) Top 19 high-paying



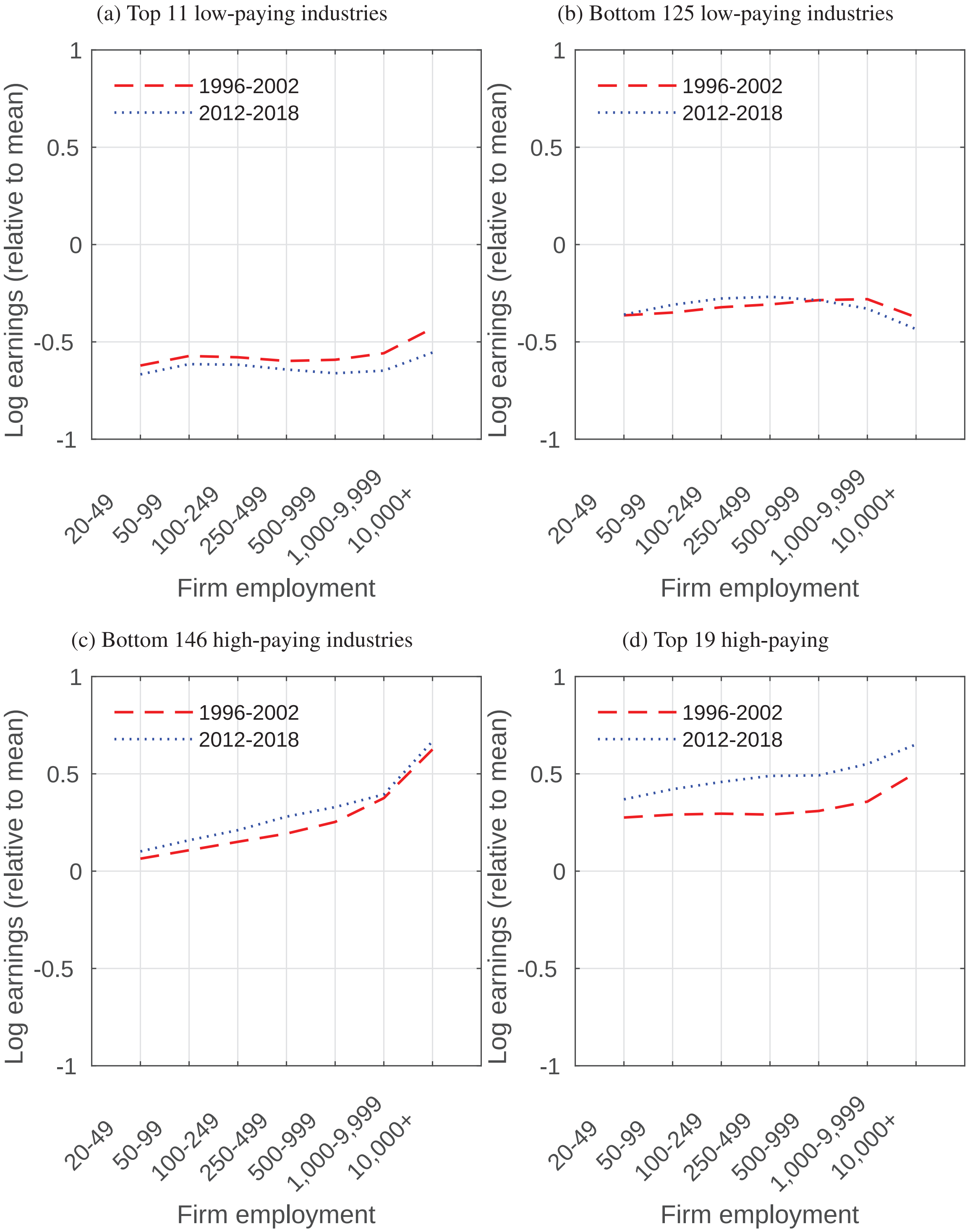
Notes: Persons with annual real earnings > $3770 in EINs with 20 or more employees.

Figure A4: Change in employment share by size class and industry group



Notes: Persons with annual real earnings > \$3770 in EINs with 20 or more employees.

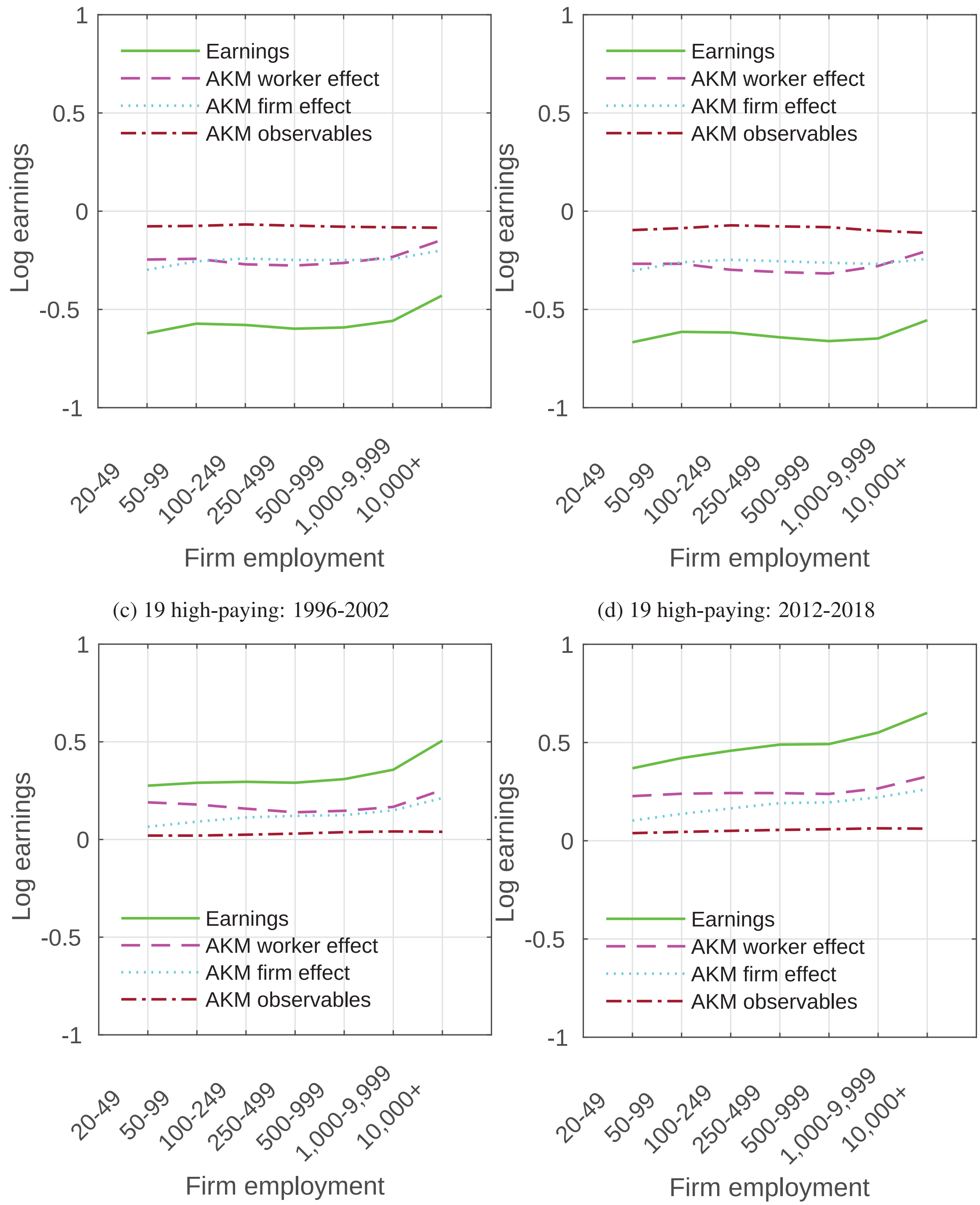
Figure A5: Earnings per worker by size class and industry group



Notes: Persons with annual real earnings > $3770 in EINs with 20 or more employees.

Figure A6: Earnings levels by size class for select industries

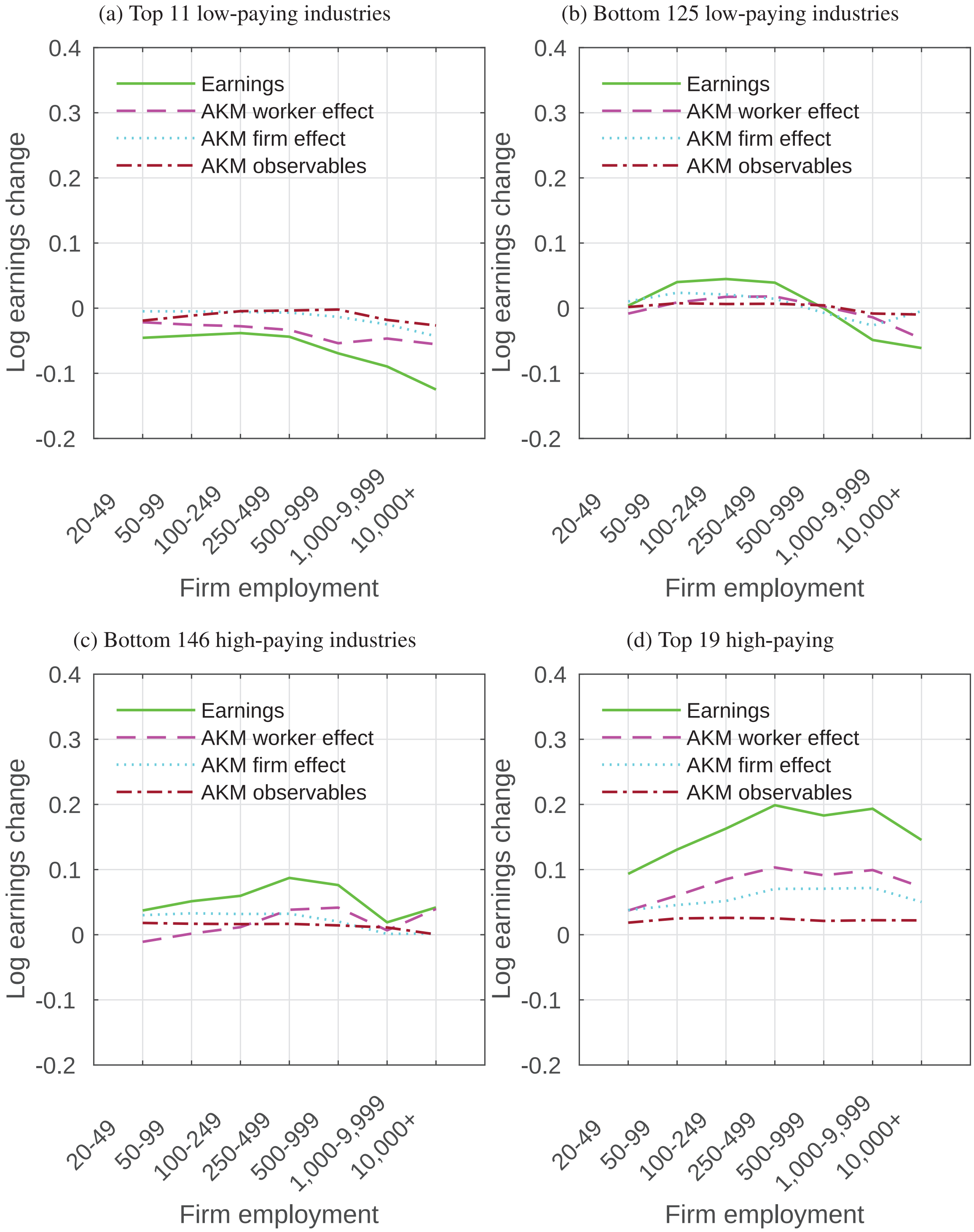
(a) 11 low-paying: 1996-2002 (b) 11 low-paying: 2012-2018



(c) 19 high-paying: 1996-2002 (d) 19 high-paying: 2012-2018

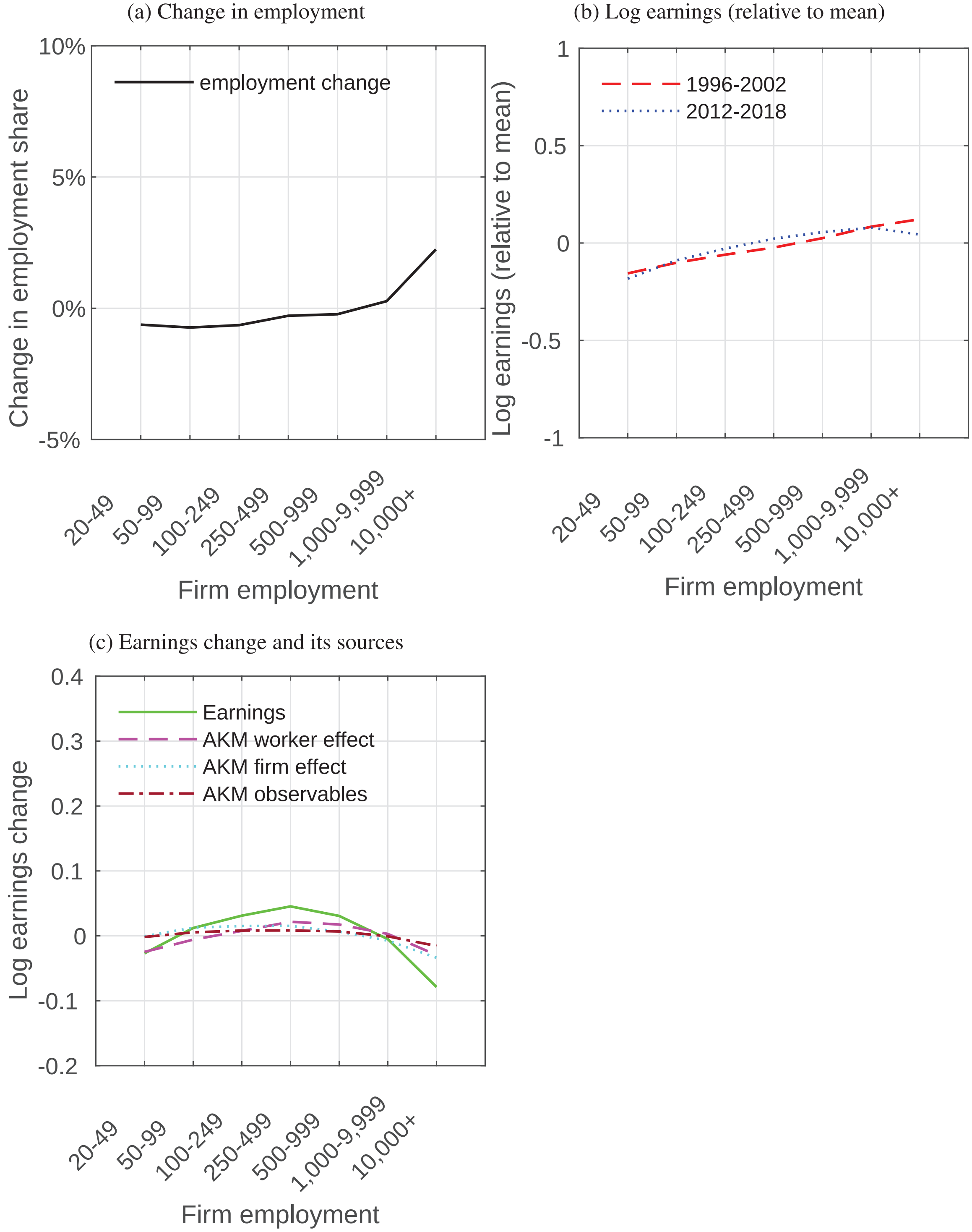
Notes: Persons with annual real earnings > $3770 in EINs with 20 or more employees.

Figure A7: Earnings change by size class for select industries, by industry group



Notes: Persons with annual real earnings > \$3770 in EINs with 20 or more employees. Size class is in terms of employment. See Equation (5) for definitions.

Figure A8: Employment and earnings by size class, national



Notes: Persons with annual real earnings > $3770 in EINs with 20 or more employees. Size class is in terms of employment.

Table A1: Variance decomposition, following Song et al. (2019)

	Interval 1: 1996-2002	Interval 2: 2004-2010	Interval 3: 2012-2018	Growth: 1 to 3
Total var($y_t^{i,j}$)	0.794	0.862	0.915	0.121
Between-firm ($\bar{y}_t^{j,k} - \bar{y}_t$)	35.4%	38.3%	42.0%	85.1%
var($\bar{\theta}^{j,k}$)	10.4%	10.8%	12.2%	23.8%
var($\psi^{j,k}$)	7.3%	8.2%	7.9%	11.6%
var($\bar{X}^{j,k}\beta$)	1.0%	0.9%	1.2%	2.7%
2cov($\bar{\theta}^{j,k}, \psi^{j,k}$)	11.7%	12.6%	13.8%	27.2%
2cov($\bar{\theta}^{j,k}, \bar{X}^{j,k}\beta$)	2.3%	2.6%	3.3%	10.3%
2cov($\bar{X}^{j,k}\beta, \psi^{j,k}$)	2.7%	3.1%	3.6%	9.5%
Within-firm var($y_t^{i,j,k} - \bar{y}_t^{j,k}$)	64.6%	61.7%	58.0%	14.9%
var($\theta^i - \bar{\theta}^{j,k}$)	42.6%	40.8%	38.5%	11.7%
var($X_t^i\beta - \bar{X}^{j,k}\beta$)	7.7%	5.8%	7.5%	6.3%
var($\varepsilon_t^{i,j,k}$)	16.1%	14.9%	13.5%	-3.5%
2cov($\theta^i - \bar{\theta}^{j,k}, X_t^i\beta - \bar{X}^{j,k}\beta$)	-2.2%	-0.1%	-1.8%	0.7%
2cov($\theta^i - \bar{\theta}^{j,k}, \varepsilon_t^{i,j,k}$)	0.2%	0.2%	0.1%	-0.2%
2cov($X_t^i\beta - \bar{X}^{j,k}\beta, \varepsilon_t^{i,j,k}$)	0.1%	0.1%	0.1%	-0.1%

Notes: Persons with annual real earnings > \$3770 in EINs with 20 or more employees. See Equation (6) for definitions.

Table A2: Variance decomposition, following Song et al. (2019), aggregated

	Interval 1: 1996-2002	Interval 2: 2004-2010	Interval 3: 2012-2018	Growth: 1 to 3
Total variance	0.794	0.862	0.915	0.121
Between-firm	35.4%	38.3%	42.0%	85.1%
Firm segregation	13.7%	14.4%	16.8%	36.8%
Firm pay premium	7.3%	8.2%	7.9%	11.6%
Firm sorting	14.4%	15.7%	17.4%	36.7%
Within-firm	64.6%	61.7%	58.0%	14.9%
Person effect	48.2%	46.5%	44.3%	18.8%
Residual	16.4%	15.2%	13.7%	-3.9%

Notes: Persons with annual real earnings > \$3770 in EINs with 20 or more employees. See Equation (6) for definitions.

Table A3: Industry-enhanced variance decomposition, in detail

	Interval 1: 1996-2002	Interval 2: 2004-2010	Interval 3: 2012-2018	Growth: 1 to 3
Total var$(y_t^{i,j,k})$	0.794	0.862	0.915	0.121
Between-firm, within-industry $\bar{y}_t^{j,k} - \bar{y}_t^k$	14.0%	14.7%	15.3%	23.1%
var$(\bar{\theta}^{j,k} - \bar{\theta}^k)$	5.2%	5.4%	5.7%	9.0%
var$(\psi^{j,k} - \bar{\psi}^k$	3.1%	3.4%	3.1%	2.9%
var$(\bar{X}^{j,k}\beta - \bar{X}^k\beta)$	0.6%	0.4%	0.6%	0.7%
2cov$[(\bar{\theta}^{j,k} - \bar{\theta}^k), (\psi^{j,k} - \bar{\psi}^k)$	3.9%	4.0%	4.3%	7.2%
2cov$(\bar{\theta}^k, \bar{X}^k\beta)$	0.6%	0.8%	0.8%	2.0%
2cov$[(\psi^{j,k} - \bar{\psi}^k), (\bar{X}^{j,k}\beta - \bar{X}^k\beta)]$	0.7%	0.7%	0.8%	1.4%
Between-industry var$(\bar{y}_t^k - \bar{y}_t)$	21.4%	23.6%	26.8%	61.9%
var$(\bar{\theta}^k)$	5.3%	5.4%	6.5%	14.8%
var$(\bar{\psi}^k)$	4.2%	4.8%	4.8%	8.7%
var$(\bar{X}^k\beta)$	0.5%	0.5%	0.7%	2.1%
2cov$(\bar{\theta}^k, \bar{\psi}^k)$	7.8%	8.6%	9.4%	19.9%
2cov$(\bar{\theta}^k, \bar{X}^k\beta)$	1.6%	1.8%	2.5%	8.3%
2cov$(\bar{\psi}^k, \bar{X}^k\beta)$	2.0%	2.4%	2.8%	8.1%
Within-firm var$(y_t^{i,j,k} - \bar{y}_t^{j,k})$	64.6%	61.7%	58.0%	14.9%
var$(\theta^i - \bar{\theta}^{j,k})$	42.6%	40.8%	38.5%	11.7%
var$(X_t^i\beta - \bar{X}^{j,k}\beta)$	7.7%	5.8%	7.5%	6.3%
var$(\varepsilon_t^{i,j,k})$	16.1%	14.9%	13.5%	-3.5%
2cov$(\theta^i - \bar{\theta}^{j,k}, X_t^i\beta - \bar{X}^{j,k}\beta)$	-2.2%	-0.1%	-1.8%	0.7%
2cov$(\theta^i - \bar{\theta}^{j,k}, \varepsilon_t^{i,j,k})$	0.2%	0.2%	0.1%	-0.2%
2cov$(X_t^i\beta - \bar{X}^{j,k}\beta, \varepsilon_t^{i,j,k})$	0.1%	0.1%	0.1%	-0.1%

Notes: Persons with annual real earnings $>$ \$3770 in EINs with 20 or more employees.

Table A4: Industry contributions to between-industry variance growth, top 10 and bottom 3 industries

4-digit NAICS	Industry title	Relative earnings: average	Relative earnings: change	Employment share: average	Employment share: change	Bet.-ind. var. growth	Bet.-ind. var. share
7225	Restaurants and Other Eating Places	-0.739	-0.027	4.9%	2.0%	0.013	16.9%
4529	Other General Merchandise Stores	-0.539	-0.051	1.4%	1.5%	0.005	6.8%
5191	Other Information Services	0.798	0.699	0.2%	0.3%	0.004	5.8%
5415	Computer Systems Design	0.663	0.012	1.7%	0.9%	0.004	5.6%
5112	Software Publishers	1.009	0.186	0.5%	0.2%	0.004	5.6%
5511	Management of Companies	0.471	0.201	2.0%	-0.1%	0.004	5.0%
4451	Grocery Stores	-0.378	-0.194	2.4%	0.0%	0.004	4.7%
6221	General Medical & Surg. Hospitals	0.205	0.170	4.5%	0.5%	0.003	4.2%
6241	Individual and Family Services	-0.490	-0.155	0.8%	0.6%	0.003	3.5%
5239	Other Financial Invest. Activities	0.834	0.388	0.3%	0.1%	0.003	3.3%
6231	Skilled Nursing Care Facilities	-0.375	0.079	1.5%	-0.1%	-0.001	-1.5%
4521	Department Stores	-0.593	-0.142	1.6%	-1.1%	-0.001	-1.5%
3341	Computer Manufacturing	0.911	0.191	0.5%	-0.4%	-0.001	-1.6%

Notes: Persons with annual real earnings > \$3770 in EINs with 20 or more employees. Average log earnings for industry k are relative to the economy average. Industry k's contribution to between-industry variance growth is in terms of Equation (3).

Table A5: Sources of industry contributions to between-industry variance growth, top 10 and bottom 3 industries

4-digit NAICS	Industry title	Bet.-ind. var. share	Segregation	Pay premia	Sorting	Shift share: earnings	Shift share: employment
7225	Restaurants and Other Eating Places	16.9%	40.3%	12.5%	47.1%	15.5%	84.5%
4529	Other General Merchandise Stores	6.8%	36.2%	16.3%	47.5%	15.0%	85.0%
5191	Other Information Services	5.8%	28.0%	22.3%	49.7%	51.5%	48.5%
5415	Computer Systems Design	5.6%	62.3%	1.7%	35.9%	6.5%	93.5%
5112	Software Publishers	5.6%	43.0%	11.4%	45.6%	45.5%	54.5%
5511	Management of Companies	5.0%	49.8%	8.4%	41.8%	103.5%	-3.5%
4451	Grocery Stores	4.7%	28.3%	21.3%	50.4%	99.5%	0.5%
6221	General Medical & Surg. Hospitals	4.2%	19.5%	28.7%	51.8%	92.9%	7.1%
6241	Individual and Family Services	3.5%	43.5%	11.5%	45.0%	45.8%	54.2%
5239	Other Financial Invest. Activities	3.3%	46.6%	10.0%	43.4%	64.4%	35.6%
6231	Skilled Nursing Care Facilities	-1.5%	39.7%	13.3%	47.0%	82.0%	18.0%
4521	Department Stores	-1.5%	24.3%	28.1%	47.6%	-242.4%	342.4%
3341	Computer Manufacturing	-1.6%	11.5%	34.8%	53.6%	-145.9%	245.9%

Notes: Notes: Persons with annual real earnings > \$3770 in EINs with 20 or more employees. Industry k's contribution to between-industry variance growth is in terms of Equation (3). The shift-share calculations follow Equation (4).

B The between-firm, within-industry component of rising earnings inequality

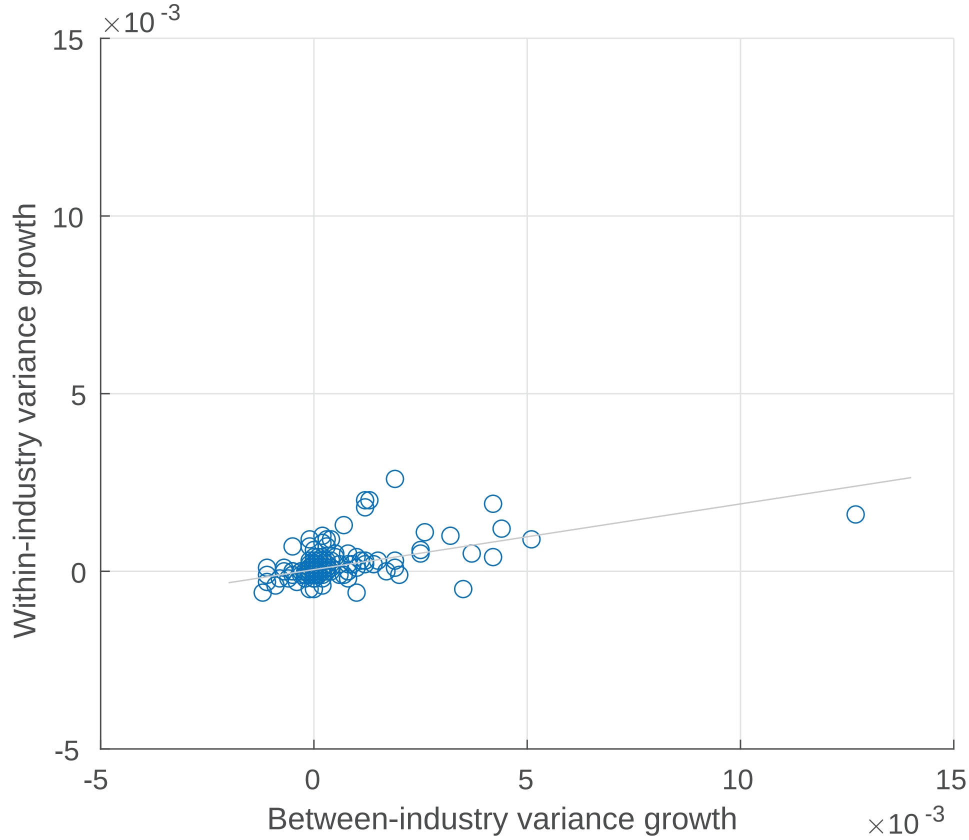
This appendix includes discussion of the between-firm, within-industry component of rising earnings inequality. Tables B4 and B5 show patterns of the top ten industries for the between-firm, within-industry contribution. The top ten industries alone contribute 65% to the between-firm, within-industry component while accounting for only 17% of employment. Four of the top ten industries in Table B4 are also among the top ten industries (for the between-industry component) in Table A4. These industries include Computer Systems Design (5415), Other Information Services (5191), Restaurants and Other Eating Places (7225), and Individual and Family Services (6241). For the six non-overlapping 4-digit industries, five overlap at the 3-digit or 2-digit level.

The overlap in the ranking of industries in terms of the between-industry component and between-firm, within-industry component is far from perfect. A good example of this is Grocery Stores which is in the bottom three for the between-firm, within-industry component (contributing negatively) and in the top ten for the between-industry component. This is a low-earnings industry that has exhibited a substantial decrease in average earnings (see Table 3) with an accompanying decrease in the firm premium (Table A4). However, within the industry, there has been a modest compression of earnings across firms within the industry. Most of this is due to decrease in sorting across firms within the industry.

While there is a strong relationship between the magnitude of the between-firm, within-industry components and the between-industry components, the between-industry components are much smaller in magnitude. This translates into a slope coefficient in Figure B1 of 0.18.

Tables B1 and B2 illustrate that the within-industry, between-firm component is also concentrated in a relatively small fraction of industries. The top 36 industries with a contribution in excess of 1% account for more than 100% of the overall within-industry, between-firm contribution. 24 of the top 36 are high-paying industries, and similar to the between-industry, high-paying industry results, earnings changes are relatively more important than employment changes in accounting for their contribution. In contrast, for the 12 low-paying industries in the top 36, employment changes are relatively more important than earnings changes.

Figure B1: Industry contributions to between-industry variance growth and within-industry variance growth



Notes: Persons with annual real earnings > $3770 in EINs with 20 or more employees.

Table B1: Industry contributions to within-industry variance growth by variance contribution and average earnings

Industry share of within-industry variance growth	Number of industries	Total employment share	Total contribution to within-industry variance growth	Total share of within-industry variance growth
$> 5\%$	6 industries	13.9%	0.012	42.6%
1% to 5%	30 industries	24.2%	0.019	68.0%
0.05% to 1%	84 industries	25.2%	0.009	31.7%
-0.05% to 0.05%	73 industries	6.0%	-0.000	-0.5%
$< -0.05\%$	108 industries	30.7%	-0.012	-41.8%
Overall	301 industries	100.0%	0.028	100.0%

Notes: Persons with annual real earnings > $3770 in EINs with 20 or more employees. Employment shares are calculated as the average of 1996-2002 and 2012-2018 employment shares.

Table B2: Industry contributions to within-industry variance growth by variance contribution and average earnings

Industry relative pay	Number of industries	Emp. share	Within-ind. var. growth	Within-ind var. cont.	Shift-share: employment	Shift-share: earnings
			36 industries with variance contribution > 1%			
High-paying	24 industries	21.7%	0.019	66.9%	44.0%	56.0%
Low-paying	12 industries	16.4%	0.012	43.7%	60.8%	39.2%
			265 industries with variance contribution ≤ 1%			
High-paying	141 industries	34.3%	-0.005	-18.1%		
Low-paying	124 industries	27.6%	0.002	7.5%		
Overall	301 industries	100.0%	0.028	100.0%	86.4%	13.6%

Notes: Persons with annual real earnings > $3770 in EINs with 20 or more employees. Employment shares are calculated as the average of 1996-2002 and 2012-2018 employment shares. Shift-share results are summed across industries and normalized by the total contribution so that the two components sum to 100%. The two rows for the 265 industries with variance contribution ≤ 1% have missing cells because the denominator for the shift-share decomposition is close to zero.

Table B3: Sources of within- industry variance growth, by top 36 industries

Industry relative earnings	Number of industries	Total contribution to within-industry variance growth	Share of contribution explained by within-industry: segregation	pay premium	sorting
High-Paying	24 industries	66.9%	47.5%	13.3%	39.2%
Low-Paying	12 industries	43.7%	46.0%	15.7%	38.3%

Notes: Persons with annual real earnings > $3770 in EINs with 20 or more employees.

Table B4: Industry contributions to within-industry variance growth, top 10 and bottom 3 industries

4-digit		Relative earnings:		Employment share:		Within-ind.	Within-ind.
NAICS	Industry title	average	change	average	change	var. growth	var. contrib.
5613	Employment Services	-0.685	0.017	3.9%	0.6%	0.003	9.4%
5416	Management & Consulting	0.381	0.069	0.9%	0.6%	0.002	7.1%
6211	Offices of Physicians	0.254	0.098	1.7%	0.5%	0.002	7.0%
5415	Computer Systems Design	0.663	0.012	1.7%	0.9%	0.002	6.9%
6216	Home Health Care Services	-0.525	-0.016	0.8%	0.4%	0.002	6.5%
7225	Restaurants & Other Eating Places	-0.739	-0.027	4.9%	2.0%	0.002	5.8%
6214	Outpatient Care Centers	0.167	0.250	0.6%	0.5%	0.001	4.6%
5191	Other Information Services	0.798	0.699	0.2%	0.3%	0.001	4.3%
6241	Individual and Family Services	-0.490	-0.155	0.8%	0.6%	0.001	4.0%
4541	Electronic Shopping & Mail-Order	0.064	0.446	0.3%	0.1%	0.001	3.7%
4451	Grocery Stores	-0.378	-0.194	2.4%	0.0%	-0.001	-1.9%
3344	Semiconductor Manufacturing	0.556	0.299	0.8%	-0.5%	-0.001	-2.0%
3341	Computer Manufacturing	0.911	0.191	0.5%	-0.5%	-0.001	-2.3%

Notes: Persons with annual real earnings > \$3770 in EINs with 20 or more employees. Average log earnings for industry k are relative to the economy average.

Table B5: Industry contributions to within-industry variance growth, top 10 and bottom 3 industries

4-digit NAICS	Industry title	Within-ind. var. share	Segregation	Pay premia	Sorting	Shift share: earnings	Shift share: employment
5613	Employment Services	9.4%	53.6%	11.0%	35.4%	54.5%	45.5%
5416	Management & Consulting	7.1%	51.5%	12.0%	36.5%	9.7%	90.3%
6211	Offices of Physicians	7.0%	50.8%	10.8%	38.5%	52.4%	47.6%
5415	Computer Systems Design	6.9%	55.7%	12.0%	32.3%	17.2%	82.8%
6216	Home Health Care Services	6.5%	38.1%	14.9%	47.0%	38.0%	62.0%
7225	Restaurants & Other Eating Places	5.8%	55.9%	13.0%	31.1%	25.9%	74.1%
6214	Outpatient Care Centers	4.6%	40.0%	14.6%	45.4%	48.0%	52.0%
5191	Other Information Services	4.3%	26.4%	29.8%	43.8%	28.3%	71.7%
6241	Individual and Family Services	4.0%	31.5%	27.9%	40.5%	27.8%	72.2%
4541	Electronic Shopping & Mail-Order	3.7%	43.7%	11.7%	44.7%	69.3%	30.7%
4451	Grocery Stores	-1.9%	27.8%	9.3%	63.0%	101.1%	-1.1%
3344	Semiconductor Manufacturing	-2.0%	16.1%	28.6%	55.4%	-128.1%	228.1%
3341	Computer Manufacturing	-2.3%	43.8%	20.3%	35.9%	19.3%	80.7%

Notes: Persons with annual real earnings > \$3770 in EINs with 20 or more employees.

C Results by gender

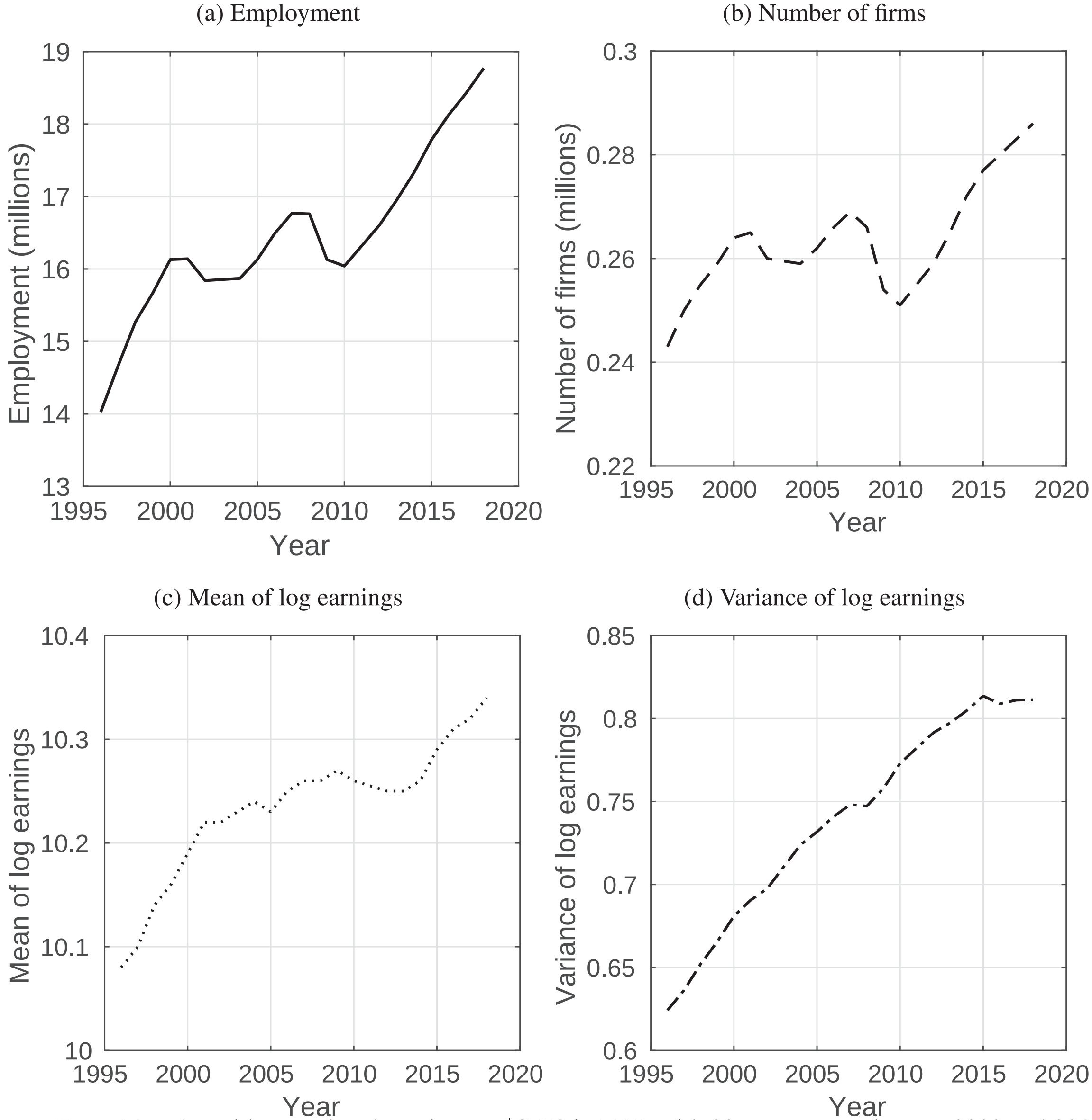
In this Appendix, we present the tables and figures with results separately by gender. The results for females and males separately are very similar both qualitatively and quantitatively with each other and with the pooled males and females results presented in the main text. For pooled males and females, females only, and males only, rising between-industry dispersion is the most important component of rising overall earnings dispersion and the rising between-industry dispersion is concentrated in a relatively small number of industries. The top ten industries accounting for rising between-industry dispersion overlap considerably with seven being the same for both females and males. The exceptions generally show up in the top ten percent of industries that account for virtually all of the increase in between-industry dispersion. Rising between-industry dispersion is quantitatively more important for males than females. However, the patterns of the relative contributions of sorting, segregation, and firm premia effects as well as the patterns of contributions in high-paying vs. low-paying industries are very similar.

Using the Song et al. (2019) results for the periods that most closely overlap with ours (1994-2000 to 2007-13), they find that 86.5% of variance growth for males is between firms, which is very similar to our result of 84.5%. Estimates of the sorting, segregation, and firm premia effects are also similar between Song et al. (2019) and our results for similar time periods. For example, for the variance growth from the mid-to-late 1990s to the most recent period, segregation contributes 35.5% in the Song et al. (2019) analysis and 37.3% in our LEHD data. Sorting contributes 37.5% in the Song et al. (2019) analysis and 35.3% in our LEHD data. Rising dispersion in firm premium contributes 14.6% in Song et al. (2019) and 11.8% in the LEHD.

Turning to females and again using the Song et al. (2019) results for the periods that most closely overlap with ours (1994-2000 to 2007-13), they find that 73.4% of females earnings variance growth is between firms, which is very similar to our result of 71.4%. Estimates of the sorting, segregation, and firm premia effects are also similar between Song et al. (2019) and our results for similar time periods. For example, for the variance growth of females from the mid-to-late 1990s to the most recent period, segregation contributes 28.7% in the Song et al. (2019) analysis and 31.2% in our LEHD data. Sorting contributes 33.0% in the Song et al. (2019) analysis and 32.0% in our LEHD data. Rising dispersion in firm premium contributes 11.7% in Song et al. (2019) and 8.3% in the LEHD.

While the findings on the respective contributions of between-firm dispersion and the components

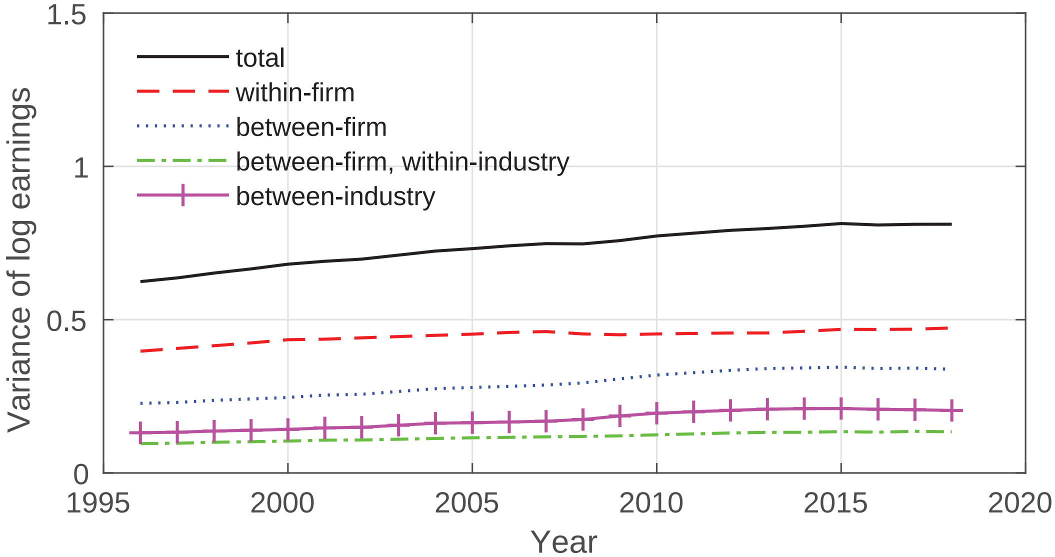
Figure C1: Descriptive statistics, females only



Notes: Females with annual real earnings $>$ \$3770 in EINs with 20 or more employees. 2003 and 2011 are linearly interpolated.

in terms sorting, segregation, and firm premia match Song et al. (2019) results closely for males and females (and in turn the pooled results presented in the main text), the key difference is that we find that these patterns reflect between-industry effects in a relatively small number of industries. For example, for females as well as males, we find it is increased sorting of industries with low (high) average person effects and low(high) average firm effects in a small number of industries such as restaurants and other eating places (software publishers) that dominates increasing earnings inequality.

Figure C2: Variance decomposition by year, 1996-2018, females only



Notes: Females with annual real earnings > $3770 in EINs with 20 or more employees. 2003 and 2011 are linearly interpolated.

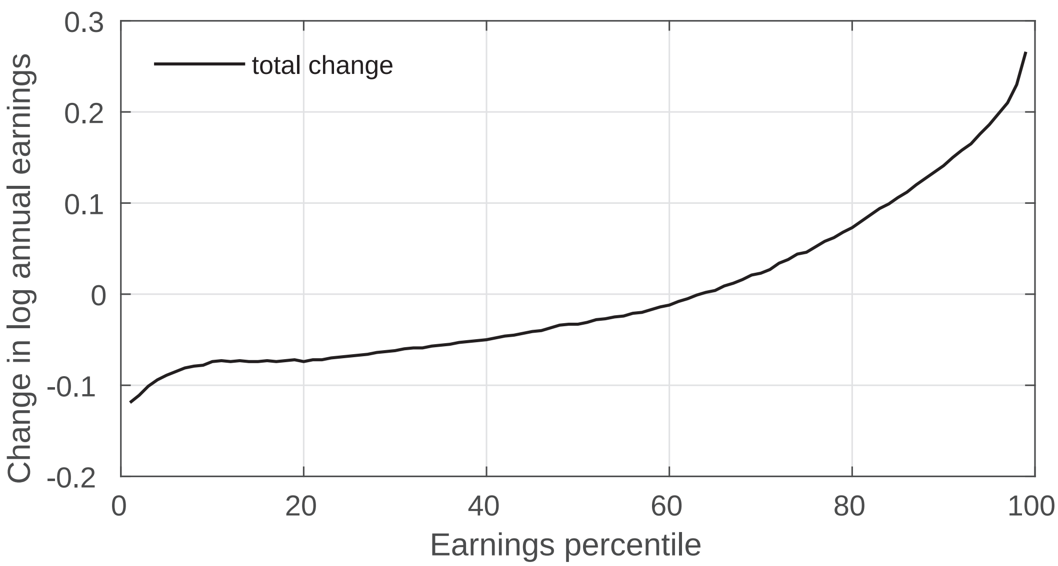
Table C1: Variance decomposition by 7-year interval, females only

	Interval 1: 1996-2002	Interval 2: 2004-2010	Interval 3: 2012-2018	Growth: 1 to 3
Variance, in levels:				
Total var$(y_t^{i,j,k} - \bar{y}_t)$	0.668	0.746	0.807	0.139
within-firm var$(y_t^{i,j,k} - \bar{y}_t^{j,k})$	0.434	0.463	0.474	0.040
Within-industry var$(\bar{y}_t^{j,k} - \bar{y}_t^{k})$	0.094	0.110	0.127	0.032
Between-industry var$(\bar{y}_t^{k} - \bar{y}_t)$	0.139	0.173	0.207	0.067
Variance, as percent of total:				
within-firm var$(y_t^{i,j,k} - \bar{y}_t^{j,k})$	65.0%	62.0%	58.7%	28.6%
Within-industry var$(\bar{y}_t^{j,k} - \bar{y}_t^{k})$	14.1%	14.8%	15.7%	23.2%
Between-industry var$(\bar{y}_t^{k} - \bar{y}_t)$	20.9%	23.2%	25.6%	48.2%
Other measures:				
Sample size (millions)	107.7	114.2	124.0	
Number of firms (thousands)	470	460	466	
Number of NAICS industries	301	301	301	

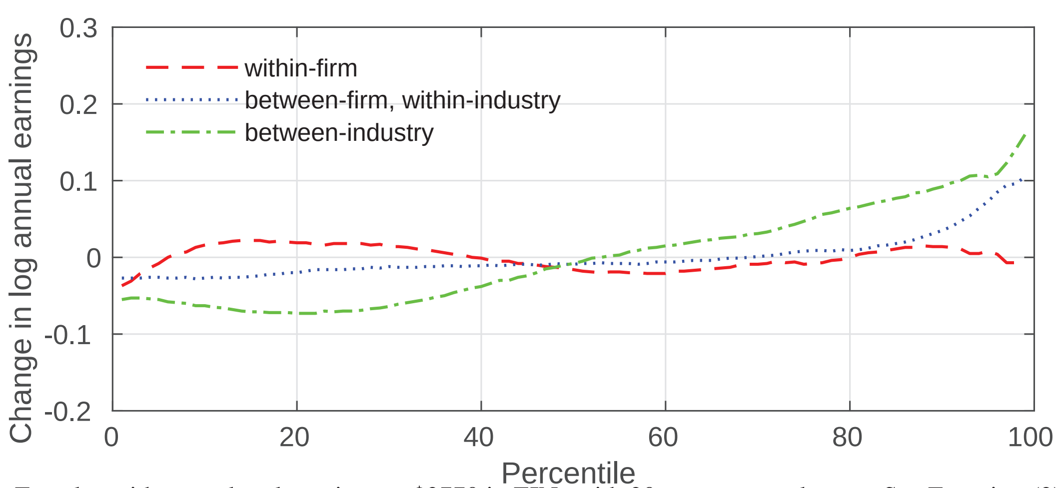
Notes: Females with annual real earnings > $3770 in EINs with 20 or more employees. See Equation (1) for definitions.

Figure C3: Change in log real annual earnings, by percentile, females only

(a) Overall change in earnings



(b) Change within and between firms and industries



Notes: Females with annual real earnings > \$3770 in EINs with 20 or more employees. See Equation (2) for definitions.

Table C2: Industry contributions to between-industry variance growth, by variance contribution, females only

Industry share of between-industry variance growth	Number of industries	Total employment share	Total contribution to between-industry variance growth	Total share of between-industry variance growth
$> 5\%$	4 industries	17.0%	0.025	37.4%
1% to 5%	23 industries	25.6%	0.039	58.4%
0.05% to 1%	76 industries	22.8%	0.015	22.3%
-0.05% to 0.05%	151 industries	12.3%	0.000	0.3%
$< -0.05\%$	47 industries	22.3%	-0.012	-18.5%
Overall	301 industries	100.0%	0.067	100.0%

Notes: Females with annual real earnings $>$ \$3770 in EINs with 20 or more employees. Employment shares are calculated as the average of 1996-2002 and 2012-2018 employment shares. See Equation (3) for definitions.

Table C3: Industry contributions to between-industry variance growth, top 27 industries, females only

4-digit NAICS	Industry title	Employment share: average	Employment share: change	Relative earnings: average	Relative earnings: change	Share of bet.-ind. variance growth
2111	Oil and Gas Extraction	0.2%	-0.0%	0.930	0.395	1.8%
3254	Pharmaceutical Manufacturing	0.5%	-0.1%	0.881	0.273	2.7%
4234	Professional Equip. Wholesaler	0.6%	-0.1%	0.549	0.202	1.6%
4451	Grocery Stores	2.6%	-0.1%	-0.336	-0.135	3.3%
4481	Clothing Stores	1.1%	-0.0%	-0.476	-0.267	4.1%
4529	Othr. Genrl. Merchandise Stores	1.7%	1.7%	-0.440	-0.025	5.5%
5112	Software Publishers	0.4%	0.1%	0.969	0.181	3.4%
5191	Other Information Services	0.2%	0.2%	0.728	0.603	4.1%
5221	Depository Credit Intermediat	3.2%	-0.5%	0.220	0.170	3.2%
5239	Other Financial Invest Activity	0.3%	0.1%	0.762	0.317	2.3%
5241	Insurance Carriers	2.3%	-0.8%	0.548	0.154	2.4%
5413	Architectur & Enginr. Services	0.8%	0.1%	0.410	0.191	2.0%
5415	Computer Systems Design	1.2%	0.5%	0.632	-0.011	2.9%
5416	Management & Scientific Serv.	0.9%	0.6%	0.388	0.059	1.8%
5417	Scientific Research Services	0.7%	-0.0%	0.713	0.272	3.8%
5511	Management of Companies	2.0%	-0.1%	0.459	0.224	6.1%
5614	Business Support Services	0.9%	0.1%	-0.304	-0.134	1.2%
5617	Services to Buildings & Dwell	0.9%	0.3%	-0.517	-0.006	1.3%
6211	Offices of Physicians	2.8%	0.8%	0.237	0.100	2.7%
6214	Outpatient Care Centers	1.0%	0.8%	0.285	0.220	3.0%
6216	Home Health Care Services	1.5%	0.7%	-0.372	-0.047	2.2%
6221	General Medical & Hospitals	7.7%	0.5%	0.359	0.128	11.6%
6233	Continuing Care Retirement	1.0%	0.6%	-0.338	-0.035	1.4%
6241	Individual and Family Services	1.3%	1.0%	-0.333	-0.181	4.2%
7139	Othr. Amusement & Recreation	0.6%	0.1%	-0.555	-0.119	1.8%
7223	Special Food Services	0.6%	0.2%	-0.487	-0.046	1.2%
7225	Restaurants & Othr. Eat Places	5.6%	2.2%	-0.637	-0.010	14.2%

Notes: Females with annual real earnings > \$3770 in EINs with 20 or more employees. Average log earnings for industry k are relative to the economy average. The 1996-2002 and 2012-2018 intervals are averaged. Changes are the growth (or decline) from 1996-2002 to 2012-2018. See Equation (3) for definitions.

Table C4: Variance decomposition, following Song et al. (2019), females only

	Interval 1: 1996-2002	Interval 2: 2004-2010	Interval 3: 2012-2018	Growth: 1 to 3
Total var$(y_t^{i,j,k} - \bar{y}_t)$	0.668	0.746	0.807	0.139
Between-firm var$(\bar{y}_t^{j,k} - \bar{y}_t)$	35.0%	38.0%	41.3%	71.4%
Firm segregation	12.7%	13.5%	15.9%	31.1%
Firm pay premium	8.7%	9.5%	8.7%	8.3%
Firm sorting	13.6%	15.0%	16.8%	32.0%
Within-firm var$(y_t^{i,j,k} - \bar{y}_t^{j,k})$	65.0%	62.0%	58.7%	28.6%
Person effect	46.6%	45.7%	43.9%	31.2%
Residual	18.4%	16.3%	14.8%	-2.6%

Notes: Females with annual real earnings > \$3770 in EINs with 20 or more employees. See Equation (6) for definitions.

Table C5: Industry enhanced variance decomposition, females only

	Interval 1: 1996-2002	Interval 2: 2004-2010	Interval 3: 2012-2018	Growth: 1 to 3
Total variance	0.668	0.746	0.807	0.139
Between-firm, within-industry	14.1%	14.8%	15.7%	23.2%
Firm segregation	6.2%	6.5%	7.2%	11.8%
Firm pay premium	3.9%	4.1%	3.7%	2.5%
Firm sorting	4.0%	4.2%	4.8%	8.9%
Between-industry	20.9%	23.2%	25.6%	48.2%
Industry segregation	6.5%	6.9%	8.7%	19.3%
Industry pay premium	4.8%	5.4%	5.0%	5.8%
Industry sorting	9.6%	10.9%	11.9%	23.1%
Within-firm	65.0%	62.0%	58.7%	28.6%
Person effect	46.6%	45.7%	43.9%	31.2%
Residual	18.4%	16.3%	14.8%	-2.6%

Notes: Females with annual real earnings > \$3770 in EINs with 20 or more employees. See Equation (7) for definitions.

Table C6: Industry contributions to between-industry variance growth, top 10 and bottom 3 industries, females only

4-digit		Relative earnings:		Employment share:		Bet.-ind.	Bet.-ind.
NAICS	Industry title	average	change	average	change	var. growth	var. share
7225	Restaurants & Other Eating Places	-0.637	-0.010	5.6%	2.2%	0.010	14.2%
6221	General Medical & Surg Hospitals	0.359	0.128	7.7%	0.5%	0.008	11.6%
5511	Management of Companies	0.459	0.224	2.0%	-0.1%	0.004	6.1%
4529	Other General Merchandise Stores	-0.440	-0.025	1.7%	1.7%	0.004	5.5%
6241	Individual and Family Services	-0.333	-0.181	1.3%	1.0%	0.003	4.2%
4481	Clothing Stores	-0.476	-0.267	1.1%	-0.0%	0.003	4.1%
5191	Other Information Services	0.728	0.603	0.2%	0.2%	0.003	4.1%
5417	Scientific Research Services	0.713	0.272	0.7%	-0.0%	0.003	3.8%
5112	Software Publishers	0.969	0.181	0.4%	0.1%	0.002	3.4%
4451	Grocery Stores	-0.336	-0.135	2.6%	-0.1%	0.002	3.3%
5179	Other Telecommunications	0.561	-0.031	0.2%	-0.3%	-0.001	-1.4%
3341	Computer Manufacturing	0.867	0.206	0.4%	-0.3%	-0.001	-1.5%
5171	Wired Telecommunications Carriers	0.597	0.025	0.9%	-0.7%	-0.002	-3.3%

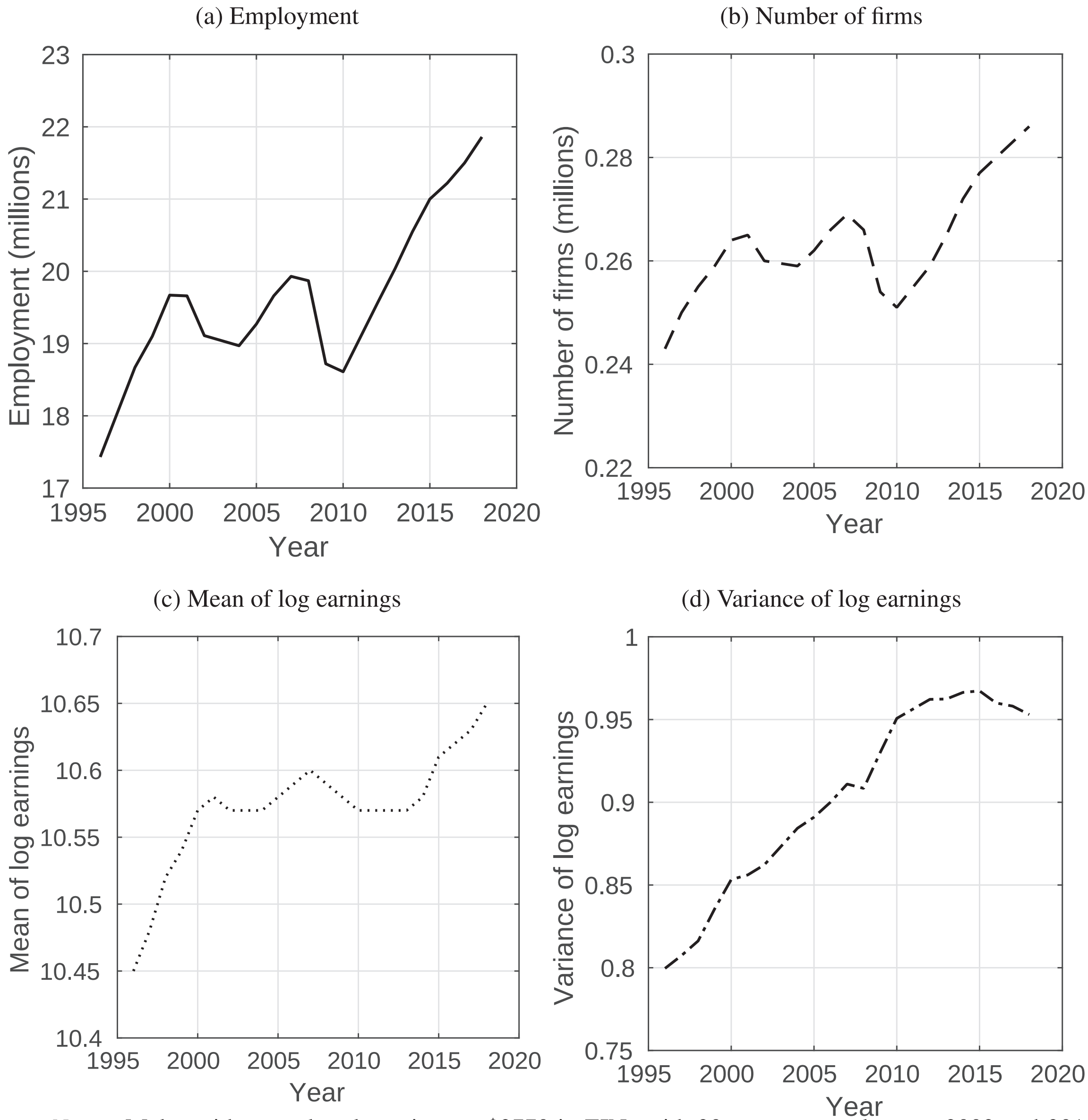
Notes: Females with annual real earnings $>$ \$3770 in EINs with 20 or more employees. Average log earnings for industry k are relative to the economy average. Industry k's contribution to between-industry variance growth is in terms of Equation (3).

Table C7: Industry contributions to between-industry variance growth, top 10 and bottom 3 industries, females only

4-digit NAICS	Industry title	Bet.-ind. var. share	Segregation	Pay premia	Sorting	Shift share: earnings	Shift share: employment
7225	Restaurants & Other Eating Places	14.2%	43.5%	9.6%	46.9%	7.6%	92.4%
6221	General Medical & Surg Hospitals	11.6%	9.7%	31.8%	58.5%	90.9%	9.1%
5511	Management of Companies	6.1%	49.7%	7.9%	42.4%	102.6%	-2.6%
4529	Other General Merchandise Stores	5.5%	34.5%	17.8%	47.7%	10.1%	89.9%
6241	Individual and Family Services	4.2%	43.6%	11.1%	45.3%	57.6%	42.4%
4481	Clothing Stores	4.1%	22.7%	26.5%	50.8%	103.8%	-3.8%
5191	Other Information Services	4.1%	26.0%	23.7%	50.2%	52.9%	47.1%
5417	Scientific Research Services	3.8%	52.8%	3.6%	43.5%	106.1%	-6.1%
5112	Software Publishers	3.4%	39.3%	13.6%	47.1%	55.4%	44.6%
4451	Grocery Stores	3.3%	25.8%	22.5%	51.7%	105.4%	-5.4%
5179	Other Telecommunications	-1.4%	13.5%	39.7%	46.8%	7.4%	92.6%
3341	Computer Manufacturing	-1.5%	7.0%	40.9%	52.1%	-130.4%	230.4%
5171	Wired Telecommunications Carriers	-3.3%	9.9%	45.3%	44.8%	-11.9%	111.9%

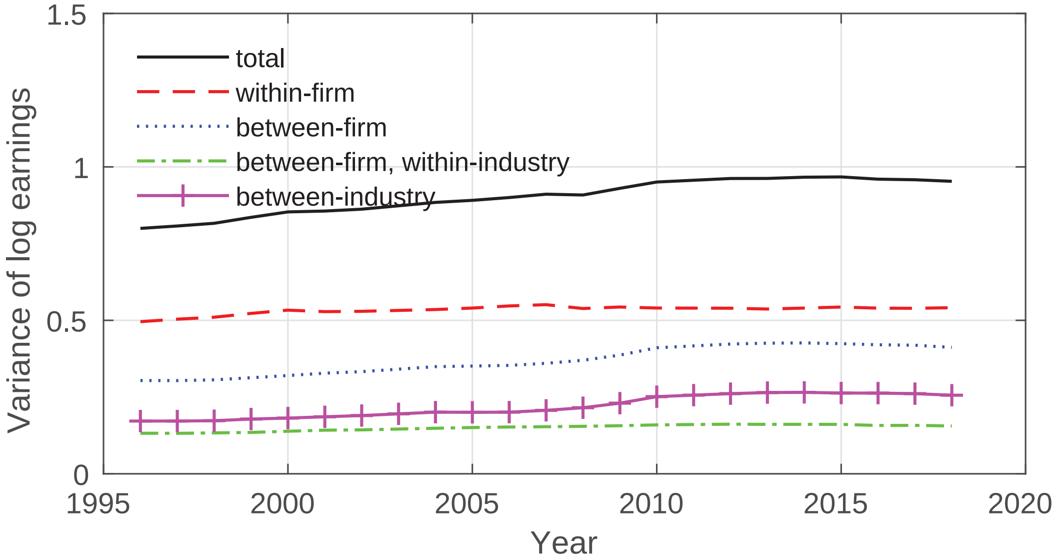
Notes: Females with annual real earnings $>$ \$3770 in EINs with 20 or more employees. Industry k's contribution to between-industry variance growth is in terms of Equation (3). The shift-share calculations follow Equation (4).

Figure C4: Descriptive statistics, males only



Notes: Males with annual real earnings > \$3770 in EINs with 20 or more employees. 2003 and 2011 are linearly interpolated.

Figure C5: Variance decomposition by year, 1996-2018, males only



Notes: Males with annual real earnings > $3770 in EINs with 20 or more employees. 2003 and 2011 are linearly interpolated.

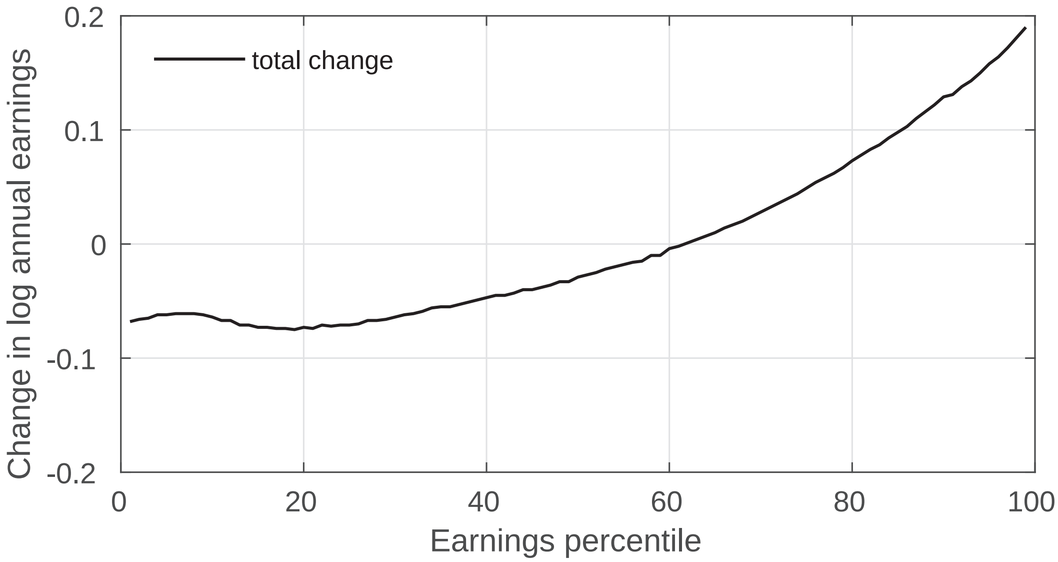
Table C8: Variance decomposition by 7-Year interval, males only

	Interval 1: 1996-2002	Interval 2: 2004-2010	Interval 3: 2012-2018	Growth: 1 to 3
Variance, in levels:				
Total variance	0.836	0.911	0.962	0.126
Within-firm	0.530	0.553	0.550	0.020
Between-firm, within-industry	0.128	0.144	0.152	0.024
Between-industry	0.178	0.214	0.261	0.083
Variance, as percent of total:				
Within-firm	63.4%	60.7%	57.1%	15.5%
Within-industry, between-firm	15.3%	15.8%	15.8%	18.9%
Between-industry	21.3%	23.5%	27.1%	65.6%
Other measures:				
Sample size (millions)	131.7	135.0	145.7	
Number of firms (thousands)	470	460	466	
Number of NAICS industries	301	301	301	

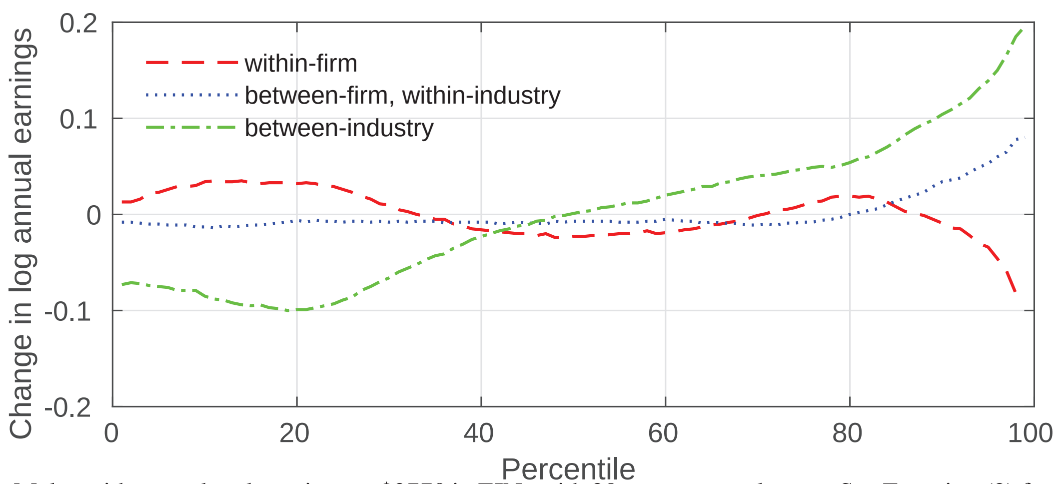
Notes: Males with annual real earnings > $3770 in EINs with 20 or more employees. See Equation (1) for definitions.

Figure C6: Change in log annual earnings by percentile, males only

(a) Overall change in earnings



(b) Change within and between firms and industries



Notes: Males with annual real earnings > \$3770 in EINs with 20 or more employees. See Equation (2) for definitions.

Table C9: Industry contributions to between-industry variance growth, by variance contribution, males only

Industry share of between-industry variance growth	Number of industries	Total employment share	Total contribution to between-industry variance growth	Total share of between-industry variance growth
$> 5\%$	7 industries	14.4%	0.049	59.0%
1% to 5%	20 industries	17.0%	0.032	38.7%
0.05% to 1%	70 industries	24.3%	0.018	21.3%
-0.05% to 0.05%	148 industries	22.2%	-0.000	-0.2%
$< -0.05\%$	56 industries	22.0%	-0.016	-18.8%
Overall	301 industries	100.0%	0.083	100.0%

Notes: Males with annual real earnings $>$ \$3770 in EINs with 20 or more employees. Employment shares are calculated as the average of 1996-2002 and 2012-2018 employment shares. See Equation (3) for definitions.

Table C10: Industry contributions to between-industry variance growth, top 27 industries, males only

4-digit NAICS	Industry title	Employment share: average	Employment share: change	Relative earnings: average	Relative earnings: change	Share of bet.-ind. variance growth
2111	Oil and Gas Extraction	0.4%	0.0%	0.961	0.195	1.8%
2131	Support Activities for Mining	0.8%	0.4%	0.237	0.212	1.3%
3241	Petroleum & Coal Manufactur	0.3%	-0.1%	0.814	0.278	1.2%
3344	Semiconductor Manufacturing	0.9%	-0.4%	0.591	0.277	1.7%
4234	Professional Equip. Wholesaler	0.9%	-0.0%	0.509	0.177	1.9%
4441	Building Material and Supplies	1.2%	0.1%	-0.372	-0.173	2.0%
4451	Grocery Stores	2.2%	0.1%	-0.390	-0.271	5.8%
4511	Sport & Hobby Stores	0.3%	0.1%	-0.674	-0.154	1.2%
4529	Othr. Genrl. Merchandise Stores	1.1%	1.3%	-0.568	-0.136	7.2%
5112	Software Publishers	0.6%	0.3%	0.966	0.174	6.2%
5182	Data Processing Services	0.3%	0.0%	0.594	0.256	1.4%
5191	Other Information Services	0.2%	0.3%	0.811	0.708	6.4%
5221	Depository Credit Intermediat.	1.2%	0.4%	0.381	0.173	2.7%
5231	Securities Brokerage	0.6%	-0.1%	0.942	0.191	1.2%
5239	Other Financial Invest. Activity	0.3%	0.2%	0.909	0.342	3.7%
5241	Insurance Carriers	1.1%	-0.1%	0.559	0.109	1.2%
5413	Architectur. & Engine Services	1.6%	0.1%	0.412	0.164	2.8%
5415	Computer Systems Design	2.2%	1.2%	0.611	0.020	6.1%
5416	Management & Scientific Serv.	0.9%	0.6%	0.387	0.054	1.6%
5417	Scientific Research Services	0.8%	-0.1%	0.739	0.235	2.9%
5511	Management of Companies	2.1%	-0.1%	0.479	0.184	4.2%
5613	Employment Services	3.7%	1.1%	-0.814	0.000	8.4%
6211	Offices of Physicians	0.8%	0.2%	0.711	0.019	1.7%
6221	General Medical & Hospitals	1.8%	0.3%	0.170	0.178	1.4%
6241	Individual and Family Services	0.3%	0.2%	-0.546	-0.142	1.3%
7139	Othr. Amusement & Recreation	0.7%	0.1%	-0.640	-0.088	1.5%
7225	Restaurants & Othr. Eat Places	4.4%	1.8%	-0.804	-0.058	18.8%

Notes: Males with annual real earnings > \$3770 in EINs with 20 or more employees. Average log earnings for industry k are relative to the economy average. The 1996-2002 and 2012-2018 intervals are averaged. Changes are the growth (or decline) from 1996-2002 to 2012-2018. See Equation (3) for definitions.

Table C11: Variance decomposition, following Song et al. (2019), aggregated, males only

	Interval 1: 1996-2002	Interval 2: 2004-2010	Interval 3: 2012-2018	Growth: 1 to 3
Total variance	0.836	0.911	0.962	0.126
Between-firm	36.6%	39.3%	42.9%	84.5%
Firm segregation	15.4%	16.3%	18.3%	37.4%
Firm pay premium	7.4%	8.3%	8.0%	11.8%
Firm sorting	13.8%	14.8%	16.6%	35.3%
Within-firm	63.4%	60.7%	57.1%	15.5%
Person effect	47.5%	45.8%	43.7%	18.6%
Residual	15.9%	14.9%	13.4%	-3.1%

Notes: Males with annual real earnings > $3770 in EINs with 20 or more employees. See Equation (6) for definitions.

Table C12: Industry-enhanced variance decomposition, males only

	Interval 1: 1996-2002	Interval 2: 2004-2010	Interval 3: 2012-2018	Growth: 1 to 3
Total variance	0.836	0.911	0.962	0.126
Between-firm, within-industry	15.3%	15.8%	15.8%	18.9%
Firm segregation	7.6%	7.8%	7.9%	9.7%
Firm pay premium	3.4%	3.7%	3.3%	2.4%
Firm sorting	4.2%	4.3%	4.6%	6.9%
Between-industry	21.3%	23.5%	27.1%	65.6%
Industry segregation	7.8%	8.4%	10.4%	27.7%
Industry pay premium	4.0%	4.6%	4.7%	9.4%
Industry sorting	9.5%	10.5%	12.0%	28.4%
Within-firm	63.4%	60.7%	57.1%	15.5%
Person effect	47.5%	45.8%	43.7%	18.6%
Residual	15.9%	14.9%	13.4%	-3.1%

Notes: Males with annual real earnings > $3770 in EINs with 20 or more employees. See Equation (6) for definitions.

Table C13: Industry contributions to between-industry variance growth, top 10 and bottom 3 industries males only

4-digit NAICS	Industry title	Relative earnings: average	Relative earnings: change	Employment share: average	Employment share: change	Bet.-ind. var. growth	Bet.-ind. var. share
7225	Restaurants & Other Eating Places	-0.804	-0.058	4.4%	1.8%	0.016	18.8%
5613	Employment Services	-0.814	0.000	3.7%	1.1%	0.007	8.4%
4529	Other General Merchandise Stores	-0.568	-0.136	1.1%	1.3%	0.006	7.2%
5191	Other Information Services	0.811	0.708	0.2%	0.3%	0.005	6.4%
5112	Software Publishers	0.966	0.174	0.6%	0.3%	0.005	6.2%
5415	Computer Systems Design	0.611	0.020	2.2%	1.2%	0.005	6.1%
4451	Grocery Stores	-0.390	-0.271	2.2%	0.1%	0.005	5.8%
5511	Management of Companies	0.479	0.184	2.1%	-0.1%	0.004	4.2%
5239	Other Financial Investment Activity	0.909	0.342	0.3%	0.2%	0.003	3.7%
5417	Scientific Research Services	0.739	0.235	0.8%	-0.1%	0.002	2.9%
3345	Navigational Instruments Manuf.	0.653	0.058	0.9%	-0.4%	-0.001	-1.1%
5616	Investigation and Security Services	-0.567	0.137	1.0%	0.2%	-0.001	-1.1%
3341	Computer Manufacturing	0.865	0.170	0.6%	-0.4%	-0.001	-1.4%

‘ *Notes*: Males with annual real earnings > \$3770 in EINs with 20 or more employees. Average log earnings for industry k are relative to the economy average. Industry k's contribution to between-industry variance growth is in terms of Equation (3).

Table C14: Industry contributions to between-industry variance growth, top 10 and bottom 3 industries males only

4-digit NAICS	Industry title	Bet.-ind. var. share	Segregation	Pay premia	Sorting	Shift share: earnings	Shift share: employment
7225	Restaurants & Other Eating Places	18.8%	38.3%	13.7%	48.0%	25.9%	74.1%
5613	Employment Services	8.4%	40.8%	13.2%	46.0%	0.0%	100.0%
4529	Other General Merchandise Stores	7.2%	36.2%	16.3%	47.5%	28.9%	71.1%
5191	Other Information Services	6.4%	27.0%	24.1%	48.9%	50.2%	49.8%
5112	Software Publishers	6.2%	43.5%	11.6%	44.9%	41.1%	58.9%
5415	Computer Systems Design	6.1%	63.2%	3.5%	33.3%	10.5%	89.5%
4451	Grocery Stores	5.8%	29.8%	20.6%	49.7%	96.4%	3.6%
5511	Management of Companies	4.2%	51.0%	7.3%	41.7%	104.3%	-4.3%
5239	Other Financial Investment Activity	3.7%	48.2%	9.5%	42.3%	51.2%	48.8%
5417	Scientific Research Services	2.9%	62.3%	1.3%	36.5%	116.3%	-16.3%
3345	Navigational Instruments Manuf.	-1.1%	14.9%	29.6%	55.4%	-73.9%	173.9%
5616	Investigation and Security Services	-1.1%	3.2%	22.9%	74.0%	171.7%	-71.7%
3341	Computer Manufacturing	-1.4%	8.8%	37.0%	54.3%	-161.5%	261.5%

Notes: Males with annual real earnings > \$3770 in EINs with 20 or more employees. Industry k's contribution to between-industry variance growth is in terms of Equation (3). The shift-share calculations follow Equation (4).

D Comparison to Song et al. (2019)

Our results in Table 5 show that firm-level segregation accounts for 36.8% (=11.6%+25.2%) of inequality growth over the 1996-2002 to 2012-2018 intervals, firm-level sorting accounts for 36.6% (=8.6%+28.0%), and the rising firm premia accounts for 11.6% (=2.9%+8.7%). Our results for males (Appendix Table C11) are similar: segregation 37.4%, sorting 35.3%, and pay premia 11.8%. These segregation, sorting, and firm premia results for males are similar to those of males in Song et al. (2019) when looking at variance growth over their 1994-2000 to 2007-2013 intervals: segregation 35.5%, sorting 37.5%, and pay premia 14.6%. These contributions are broadly similar to those in the longer time interval (1980-1986 to 2007-2013) reported in Song et al. (2019) with one notable exception: there is a smaller role for firm premia in the longer time interval (-1.4%). We also find a close correspondence of results for females to those reported in Song et al. (2019) over similar time periods.[1] These results imply that our findings indicate that the between-firm contribution to increasing inequality reported by Song et al. (2019) is largely – but not entirely – determined at the industry level.

The tight relationship of the role of between-firm effects comparing our results with Song et al. (2019) mitigates concerns about our use of an 18-state LEHD sample. Any such concerns are further mitigated by the analysis we report from the LBD in Appendix Table F1. In our 18-state LEHD results, 73% of the rising between firm dispersion is accounted for by rising between-industry dispersion. In the 18-state and 50-state LBD, the analogous statistics are 74% and 73%, respectively. Taken together, the comparisons with Song et al. (2019) and calculations from the LBD imply that the patterns we report here from administrative data are robust to using an 18-state or 50-state sample.

E Occupation by industry dataset and analysis

We downloaded the May OES national industry specific data from https://www.bls.gov/oes/tables.htm. We start our analysis in 2002, as the 1996-2001 are published on a SIC basis. The unit of observation in each year-specific file is 4-digit industry and occupation. We restrict the data in each year to keep only industry totals and major (2 digit) occupations. We end our analysis in 2016 because 2017 and

[1] Compare our results in Appendix Table C4 with the Song et al. (2019) results in their Appendix Table A.9.

2018 have a different level of industry aggregation.[2] There are 22 major occupations.[3]

We then create a balanced panel of 284 industries for each year 2002-2016.[4] There are nine industries that are published by OES for some years 2002-2016 but not all.[5] We had to adjust our analysis dataset because OES switched to NAICS 2012 in 2012. One of the biggest changes with the NAICS 2012 was replacing NAICS 7221 "Full Service Restaurants" and 7222 "Limited Service Eating Places" with adding NAICS 7225 "Restaurants and Other Eating Places."[6]

We use OES data to derive a balanced panel of 284 industries for each year 2002-2016. For reasons stated above, we do not include 2017 and 2018 in creating the balanced panel. The unit of observation is an industry-year. 284 industries by 15 years yields a total of 4260 observations.

We now add one more variable into the OES balanced panel: an indicator of whether the 4-digit NAICS industry is in HHS's 19 high-paying industries (H19), is in HHS's 11 low-paying industries (L11), is in HHS's 146 other high-paying industries (H146), or is in HHS's 125 other low-paying industries (L125).

Merging the balanced panel OES (284 industries) with the 301 industries in HHS is not a one-to-one merge. There are 20 industries in HHS that are not in the OES, which are mostly in agriculture.[7] There are 3 industries in the OES that are not in HHS.[8] The OES–HHS linked data is therefore a balanced panel of 281 industries for each year 2002-2016. The unit of observation is an industry-year

[2]For example, published OES industry 3250A1 in 2017 aggregates industries 3251, 3253, 3253, and 3259, which are published in detail in earlier years.

[3]These are: Management (11), Business and Financial Operations (13), Computer and Mathematical Science (15), Architecture and Engineering (17), Life, Physical, and Social Science (19), Community and Social Services (21), Legal (23), Education, Training, and Library (25), Arts, Design, Entertainment, Sports, and Media (27), Healthcare Practitioner and Technical (29), Healthcare Support (31), Protective Service (33), Food Preparation and Serving Related (35), Building and Grounds Cleaning and Maintenance (37), Personal Care and Service (39), Sales and Related (41), Office and Administrative Support (43), Farming, Fishing, and Forestry (45), Construction and Extraction (47), Installation, Maintenance, and Repair (49), Production (51), Transportation and Material Moving (53).

[4]We do not include 2017 and 2018, as including these two years would result in a balanced panel of only 225 industries.

[5]These are Postal Service (4911), Internet Publishing (5161), Telecommunications Carriers (5173), Satellite Telecommunications (5174), Cable Distribution (5175) Internet Service Providers (5181), Monetary Auth Central Bank (5211), Insurance Benefit Funds (5251), and Other Investment Funds (5259).

[6]In 2011, NAICS 7221 and 7222 had 4.6 and 4.1 million employees, respectively, and 91% of employment in each of these industries was in occupation 35 "Food Preparation and Serving." In 2012, NAICS 7225 had 8.9 million employees, with 91% in occupation 35. We have recoded all occurrences of NAICS 7221 and 7222 in 2002-2011 to NAICS 7225.

[7]Specifically, these are Oilseed and Grain Farming (1111), Vegetable and Melon Farming (1112), Fruit and Tree Nut Farming (1113), Greenhouse Nursery (1114), Other Crop Farming (1119), Cattle Ranching and Farming (1121), Hog and Pig Farming (1122), Poultry and Egg Production (1123), Aquaculture (1125), Other Animal Production (1129), Timber Tract Operations (1131), Forest Nurseries (1132), Fishing (1141), Support Activities Forestry (1153), Postal Service (4911), Satellite Telecommunications (5174), Monetary Auth. Central Bank (5211), Insurance Benefit Funds (5251), Other Investment Funds (5259), and Private Households (8141).

[8]These are Federal Executive Branch and US Postal Service (9991), State Government (9992), and Local Government (9993).

(281 industries * 15 years = 4215 observations).

The results in Table 8 are based on estimating the following equation for the sub-intervals 2002-03 and 2015-16 (for occupations indexed by j and industries indexed by k, estimated on employment-weighted basis):

$$y_{j,k} = Occupation_{j,k}\beta_2 + Industry_{j,k}\beta_3 + \varepsilon_{j,k}. \tag{A1}$$

Taking (employment-weighted) variances of both sides of equation (A1) yields: [9]

$$\begin{aligned}\underbrace{\text{var}(y_{j,k})}_{\substack{\text{earnings}\\ \text{variance}}} = & \underbrace{\text{var}(Occupation_{j,k}\beta_2 - \overline{Occupation_k\beta_2})}_{\substack{\text{within-industry dispersion}\\ \text{from occupation}}} + \underbrace{\text{var}(\overline{Occupation_k\beta_2})}_{\substack{\text{between-industry}\\ \text{segregation}}} + \\ & \underbrace{\text{var}(Industry_{j,k}\beta_3)}_{\substack{\text{between-industry}\\ \text{pay premia}}} + \underbrace{2\text{cov}(\overline{Occupation_k\beta_2}, Industry_{j,k}\beta_3)}_{\text{between-industry sorting}} + \underbrace{\text{var}(\varepsilon_{j,k})}_{\substack{\text{residual dispersion}\\ \text{(within-industry)}}}\end{aligned} \tag{A2}$$

All 19 of the industries in the HHS "19 High-Paying Industries" are in the OES balanced panel of 281 industries, and all 11 of the industries in the HHS "11 Low-Paying Industries" are in the OES balanced panel of 281 industries. 141 of the 146 "Other High-Paying Industries" and 110 of the 125 "Other Low Paying Industries" are in the OES balanced panel of 281 industries. In Table 8, we focus on the decomposition in equation (A2) for the the 19 high-paying and the 11 low-paying industries.

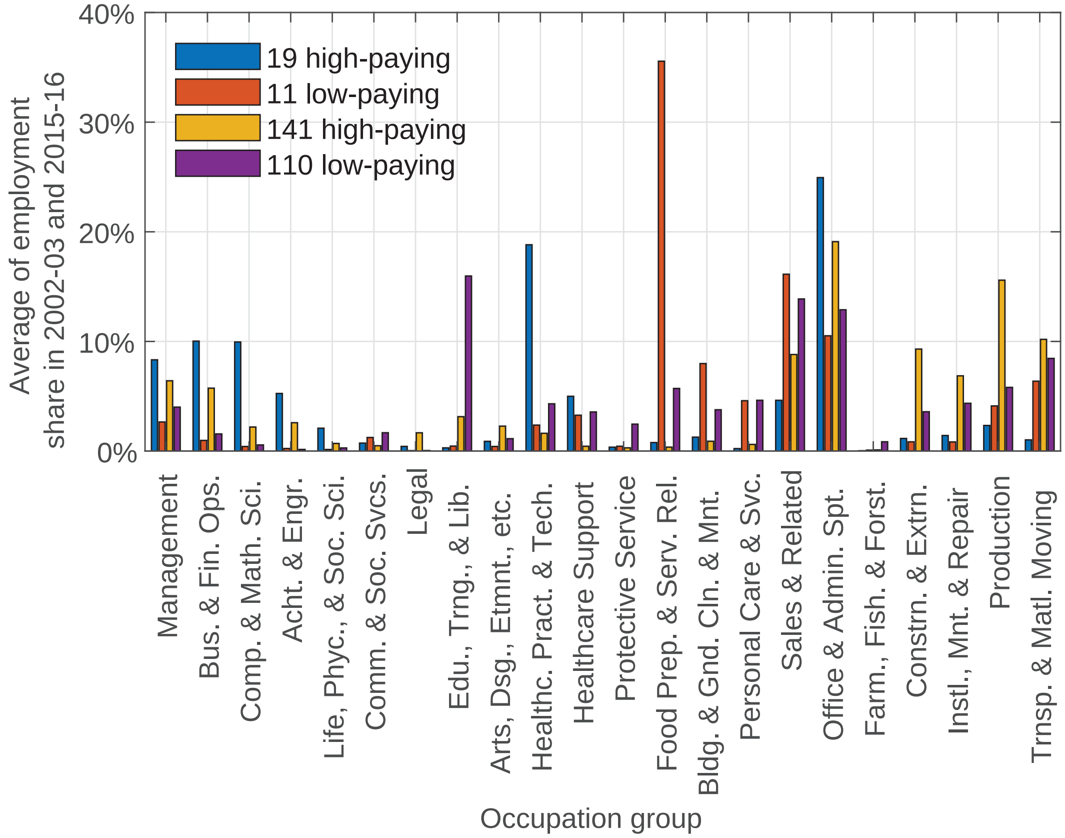
Aggregating the OES–HHS data over NAICS industries and HHS groups {19 high-paying, 11 low-paying, 146 high-paying, 125 low-paying} allows for analysis of occupational distribution in our 4 groups of industries, as shown in Appendix Figures E1 and E2, as well as Appendix Tables E1 and E2.

To explore the concentration of occupations in industries, we conduct additional exercises. First, we use the data underlying our 4-digit industry by 2-digit occupation analysis for 2002, 2003, 2015, 2016. An attractive feature of this data is that missingness is not a substantial problem and industries and occupations are harmonized across time.

Figure E3 uses our core data to quantify the share of employment in each occupation in the top 20 4-digit industries. Recall in interpreting these findings that there 281 industries in this dataset. For the

[9] The notation convention for equation (A2) is broadly similar to that for equation (7) where for example $\overline{Occupation_k\beta_2}$ is the industry mean of $Occupation_{j,k}\beta_2$. In the current context, the employment-weighting plays a critical role. The employment-weighted estimation of equation (A1) yields vectors of occupation and industry effects. The employment-weighting in the implementation of (A2) is what yields the variation within and across industries in the contribution of occupation effects.

Figure E1: Average employment by occupation group



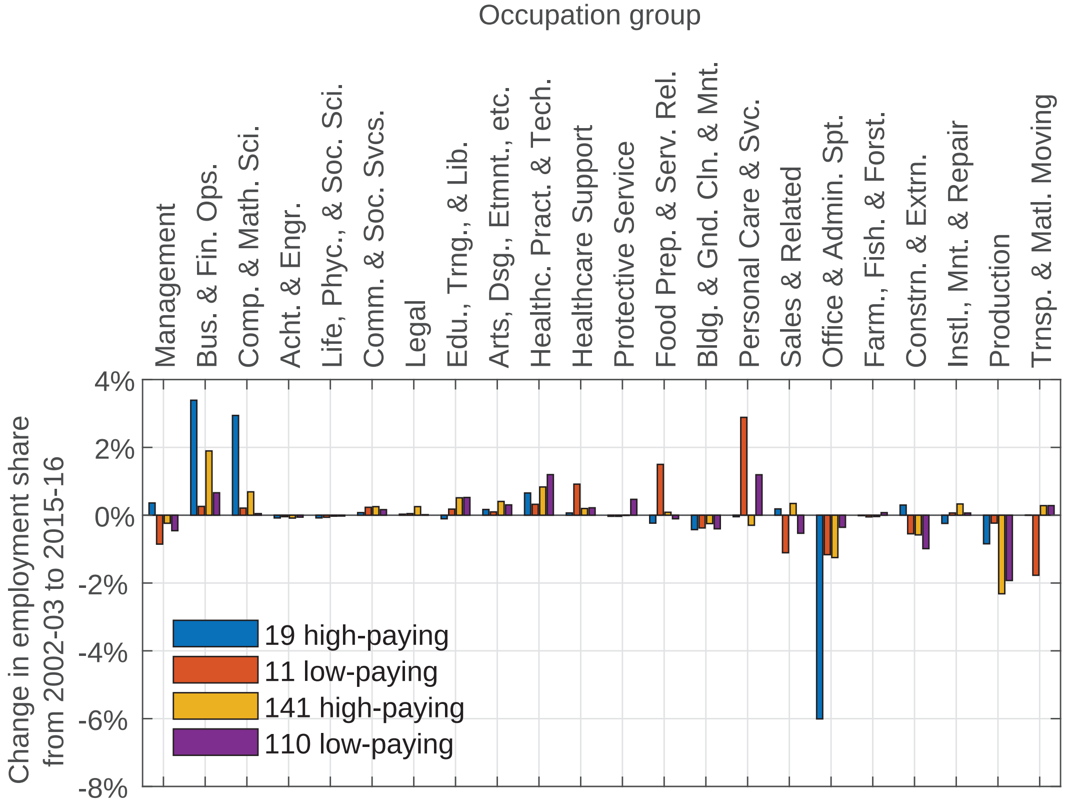
Notes: Authors' calculations of published OEWS aggregates.

median occupation, the concentration ratio is 83%. It is lowest in occupations such as Management, Office and Administrative Support, Installation and Maintenance, Production, and Transportation and Materials Moving.

Concentration of occupations is, not surprisingly, even more pronounced for more detailed occupations. For this purpose, we use the OEWS in 2015. We only use one year to avoid the concordance issues (we only constructed a harmonized dataset from 2002-2016 at the 2-digit SOC level). There are issues with missingness but this will tend to be in cells with smaller number of firms and employment. Given that the next chart examines concentration for 3-digit and 6-digit occupations, only medians are reported.

The results of this exercise are shown in Figure E4. At the 3-digit occupation level, using the top 20 industries the median is 93% and the top 4 industries the median is 59%. At the 6-digit occupation level, the top 20 industries the median is 1 and for the top 4 the median is 85%. Even for truck drivers there is substantial concentration. At the 6-digit level, 81% of tractor trailer drivers are concentrated

Figure E2: Change in employment by occupation group



Notes: Authors' calculations of published OEWS aggregates.

in the top 20 4-digit industries and 59% of tractor trailer drivers are concentrated in the top 4 4-digit industries: Cement and Concrete Product Manufacturing (3273), Grocery and Related Product Merchant Wholesalers (4244), General Freight Trucking (4841), and Specialized Freight Trucking (4842).

Table E1: Average employment by occupation and industry

	19 high-paying	11 low-paying	Other high-paying	Other low-paying
Management	8.3%	2.7%	6.4%	4.0%
Business and Financial Operations	10.0%	1.0%	5.7%	1.6%
Computer and Mathematical Science	9.9%	0.4%	2.2%	0.6%
Architecture and Engineering	5.3%	0.2%	2.6%	0.1%
Life, Physical, and Social Science	2.1%	0.1%	0.7%	0.3%
Community and Social Services	0.7%	1.2%	0.5%	1.7%
Legal	0.4%	0.0%	1.7%	0.0%
Education, Training, and Library	0.3%	0.5%	3.1%	16.0%
Arts, Design, Entertainment, Sports, and Media	0.9%	0.4%	2.3%	1.1%
Healthcare Practitioner and Technical	18.8%	2.4%	1.6%	4.3%
Healthcare Support	5.0%	3.3%	0.4%	3.6%
Protective Service	0.4%	0.4%	0.3%	2.5%
Food Preparation and Serving Related	0.8%	35.6%	0.4%	5.7%
Building and Grounds Cleaning and Maintenance	1.3%	8.0%	0.9%	3.8%
Personal Care and Service	0.2%	4.6%	0.6%	4.6%
Sales and Related	4.6%	16.1%	8.8%	13.9%
Office and Administrative Support	24.9%	10.5%	19.1%	12.9%
Farming, Fishing, and Forestry	0.0%	0.1%	0.1%	0.8%
Construction and Extraction	1.2%	0.8%	9.3%	3.6%
Installation, Maintenance, and Repair	1.4%	0.8%	6.9%	4.4%
Production	2.3%	4.1%	15.6%	5.8%
Transportation and Material Moving	1.0%	6.4%	10.2%	8.5%
OES suppressed employment	0.1%	0.3%	0.6%	0.4%

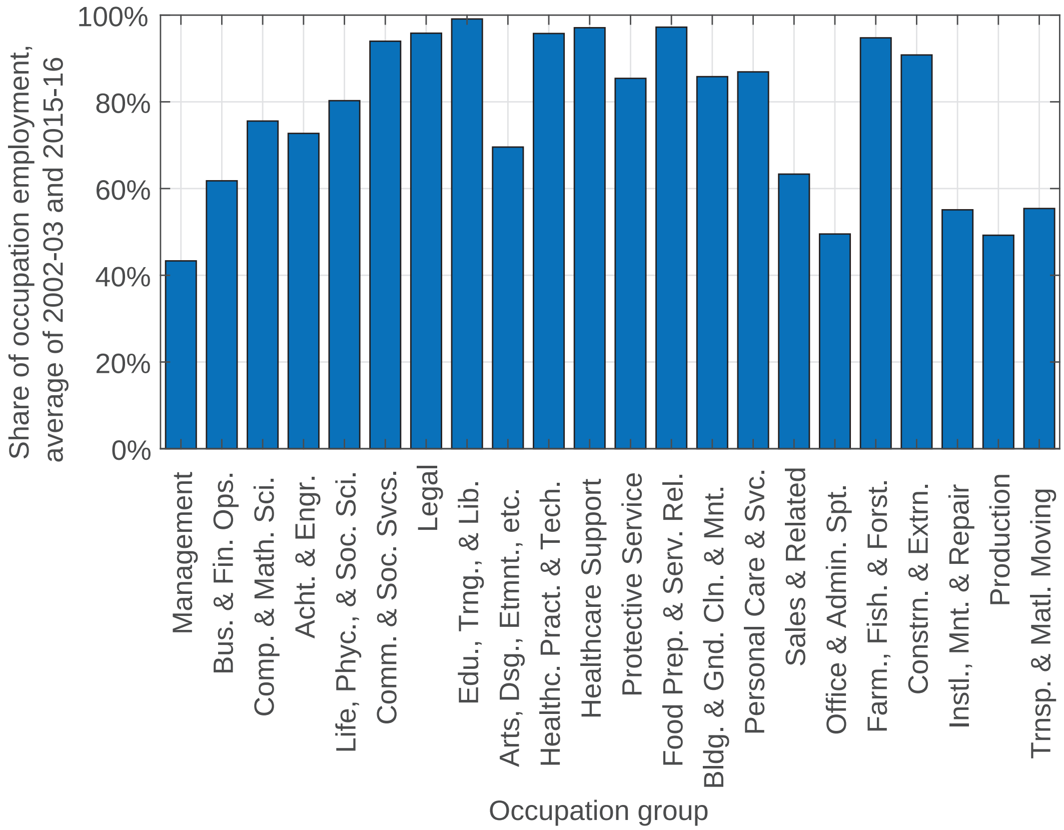
Notes: Authors' calculations of OEWS data.

Table E2: Change in employment by occupation and industry

	19 high-paying	11 low-paying	Other high-paying	Other low-paying
Management	0.4%	-0.9%	-0.2%	-0.5%
Business and Financial Operations	3.4%	0.3%	1.9%	0.7%
Computer and Mathematical Science	2.9%	0.2%	0.7%	0.0%
Architecture and Engineering	-0.1%	0.0%	-0.1%	-0.1%
Life, Physical, and Social Science	-0.1%	-0.1%	0.0%	0.0%
Community and Social Services	0.1%	0.2%	0.3%	0.2%
Legal	0.0%	0.0%	0.3%	0.0%
Education, Training, and Library	-0.1%	0.2%	0.5%	0.5%
Arts, Design, Entertainment, Sports, and Media	0.2%	0.1%	0.4%	0.3%
Healthcare Practitioner and Technical	0.7%	0.3%	0.8%	1.2%
Healthcare Support	0.1%	0.9%	0.2%	0.2%
Protective Service	0.0%	0.0%	0.0%	0.5%
Food Preparation and Serving Related	-0.2%	1.5%	0.1%	-0.1%
Building and Grounds Cleaning and Maintenance	-0.4%	-0.4%	-0.2%	-0.4%
Personal Care and Service	0.0%	2.9%	-0.3%	1.2%
Sales and Related	0.2%	-1.1%	0.3%	-0.5%
Office and Administrative Support	-6.0%	-1.2%	-1.2%	-0.4%
Farming, Fishing, and Forestry	0.0%	0.0%	0.0%	0.1%
Construction and Extraction	0.3%	-0.5%	-0.6%	-1.0%
Installation, Maintenance, and Repair	-0.2%	0.1%	0.3%	0.1%
Production	-0.8%	-0.2%	-2.3%	-1.9%
Transportation and Material Moving	0.0%	-1.8%	0.3%	0.3%
OES suppressed employment	-0.1%	-0.5%	-1.0%	-0.4%

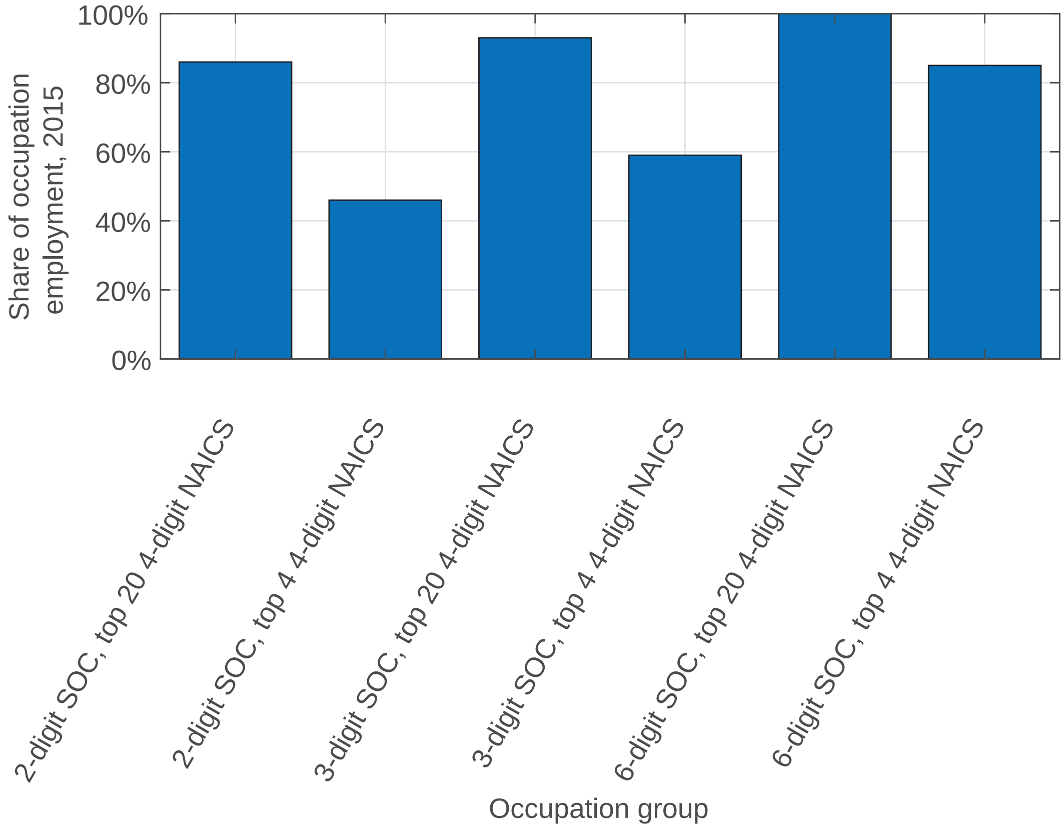
Notes: Authors' calculations of OEWS data.

Figure E3: 2-digit Occupational Employment Share, Top 20 4-digit Industries Average (2002-2003 2015-2016)



Notes: Authors' calculations of published OEWS aggregates.

Figure E4: Occupational Concentration in Industries, Median Occupation, 2015



Notes: Authors' calculations of published OEWS aggregates.

F Estimates from the Longitudinal Business Database

We use microdata from the Longitudinal Business Database (LBD) to implement a decomposition of the variance of real log earnings per worker in two sub-intervals: 1996-02 and 2012-18. For each sample, the overall between business dispersion is reported for each of these intervals and the between 4-digit industry dispersion in the variance. This is considered for various business definitions (establishment and EIN).

In all cases we start with establishment level data for each year and compute log real earnings per worker (using the public domain personal consumption deflator). Then aggregation proceeds from there depending on exercise. For establishment based, we compute total between establishment variances and then between 4-digit NAICS. We do this for establishments in all 50 states and then for a restricted sample of 18 states. For the EIN level analysis selection of state and industry is based on the dominant state and industry. Note to make results more comparable that the 18 state establishment sample is the same as the 18 state EIN sample (that is states selected if they are the EIN dominant state). Also, for the EIN 20+ employment samples, this is based on national employment for the EIN.

Results are shown in Table F1. The between-establishment, 50 state estimates indicate an industry share of between-employer earnings dispersion growth of 68.5%. Every other estimate for the industry share of growth in dispersion is between 72% and 77%. This compares to the 72.8% $(= 61.9\%/(61.9\% + 23.1\%))$ calculated from Table 1 of this paper.

Table F1: Between-employer and between-industry variance

Interval	Between-employer	Between-industry	Industry share
Establishment-level, 50 state			
1996-02	0.617	0.280	45.4%
2012-18	0.709	0.343	48.4%
Growth	0.092	0.063	68.5%
Establishment-level, 18 state			
1996-02	0.618	0.282	45.6%
2012-18	0.730	0.363	49.7%
Growth	0.112	0.081	72.3%
EIN-level, 50 State			
1996-02	0.544	0.268	49.3%
2012-18	0.627	0.331	52.8%
Growth	0.083	0.063	75.9%
EIN-level, 18 State			
1996-02	0.540	0.270	50.0%
2012-18	0.643	0.349	54.3%
Growth	0.103	0.079	76.7%
EIN-level, size 20+, 50 State			
1996-02	0.503	0.282	56.1%
2012-18	0.603	0.354	58.7%
Growth	0.100	0.072	72.0%
EIN-level, size 20+, 18 State			
1996-02	0.500	0.284	56.8%
2012-18	0.624	0.376	60.3%
Growth	0.124	0.092	74.2%

Notes: Authors' calculations of LBD data.

G Decomposition using the Human Capital Approach

We estimate the human capital earnings equation used by HLL (details of measurement and estimation in Haltiwanger, Hyatt and Spletzer (2022)) as

$$y_{i,k} = Z_{i,k}\beta_Z + Industry_{i,k}\beta_3 + \varepsilon_{i,k}, \tag{A3}$$

where Z concatenates the $AgeEduc_i$ and $Occupation_i$ vectors, and β_Z concatenates the marginal effects vectors β_1 and β_2. Define $\overline{Z_k\beta_Z}$ as the industry mean of $Z_{i,k}\beta_Z$. Taking variances of both sides of the human capital earnings equation results in:

$$\begin{aligned} \underbrace{\mathrm{var}(y_{i,k})}_{\substack{\text{earnings}\\ \text{variance}}} = \underbrace{\mathrm{var}(Z_{i,k}\beta_Z - \overline{Z_k\beta_Z})}_{\substack{\text{within-industry dispersion}\\ \text{from age, education,}\\ \text{and occupation}}} + \underbrace{\mathrm{var}(\overline{Z_k\beta_Z})}_{\substack{\text{between-industry}\\ \text{segregation}}} + \\ \underbrace{\mathrm{var}(Industry_{i,k}\beta_3)}_{\substack{\text{between-industry}\\ \text{pay premia}}} + \underbrace{2\mathrm{cov}(\overline{Z_k\beta_Z}, Industry_{i,k}\beta_3)}_{\text{between-industry sorting}} + \underbrace{\mathrm{var}(\varepsilon_{i,k})}_{\substack{\text{residual dispersion}\\ \text{(within-industry)}}} \end{aligned} \tag{A4}$$

The terms in the right hand side correspond to an alternative but analogous decomposition to that using the AKM decomposition of earnings.