

NBER WORKING PAPER SERIES

BETWEEN FIRM CHANGES IN EARNINGS INEQUALITY:
THE DOMINANT ROLE OF INDUSTRY EFFECTS

John C. Haltiwanger
James R. Spletzer

Working Paper 26786
<http://www.nber.org/papers/w26786>

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
February 2020

Any opinions and conclusions expressed herein are those of the authors and do not necessarily 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 (DRB-B0023-CED-20181213). We thank Nick Bloom and participants of an ASSA 2019 session on rising earnings inequality for helpful comments. John Haltiwanger was also a Schedule A employee of the U.S. Census Bureau at the time of the writing of this paper.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2020 by John C. Haltiwanger and James R. Spletzer. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Between Firm Changes in Earnings Inequality: The Dominant Role of Industry Effects
John C. Haltiwanger and James R. Spletzer
NBER Working Paper No. 26786
February 2020
JEL No. E24,J24,J31,L22

ABSTRACT

We find that most of the rising between firm earnings inequality that dominates the overall increase in inequality in the U.S. is accounted for by industry effects. These industry effects stem from rising inter-industry earnings differentials and not from changing distribution of employment across industries. We also find the rising inter-industry earnings differentials are almost completely accounted for by occupation effects. These results link together the key findings from separate components of the recent literature: one focuses on firm effects and the other on occupation effects. The link via industry effects challenges conventional wisdom.

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

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

I. Introduction

Recent findings emphasize that much of rise in earnings inequality is accounted for by rising between firm inequality (see, Song et. al. (2019), Barth et. al. (2016)). This follows an earlier literature emphasizing the importance of rising between firm effects for earnings inequality (e.g., Dunne, Foster and Haltiwanger (2004), Davis and Haltiwanger (1991)). Rising between firm inequality might arise from many factors. It might reflect changing quasi-rents either from changing worker shares or changing productivity and profitability dispersion. It might reflect changes in the skill mix and skill prices in the labor market that manifest themselves unevenly across firms. For example, changes in technology and globalization can change the structure of production along with the relative demand for skills.

We provide evidence that most (about 75%) of the rising between firm inequality is from rising inequality across detailed 4-digit NAICS industries. Most of this rise in inequality across industries is due to rising inter-industry earnings differentials and not due to changing distribution of employment across industries. We also find that the rising inter-industry earnings differentials are almost completely accounted for by occupation effects. For the raw rise in rising inter-industry earnings differentials, we find that about 92% of the increase is accounted for by occupation effects. About 66% of this is due to changing occupation mix across industries and 27% due to changing occupation mix differentials given the existing differences in the occupation mix across industries.

Even after we control for detailed firm-level controls, 48% of the overall rise in between firm inequality is due to rising inequality across detailed industries. About 94% of this residual rise in inter-industry earnings differentials is accounted for by occupation effects. In this case, about 26% of this is due to changing occupation mix across industries and 68% due to changing

occupation mix differentials given the existing differences in the occupation mix across industries.

Our findings build on and bring together two core components in the recent literature on rising earnings inequality. We link together the dominant role of between firm effects (e.g., Song et. al. (2019)) with the increasingly dominant role of the change in occupational mix and differentials (e.g. in the handbook paper in Acemoglu and Autor (2011)). However, the link is via industry effects which these seminal papers argued are unimportant.¹ The finding that industry-level effects play a dominant role is important because it implies that the sources of rising earnings inequality are working through channels involving structural change at the industry-level.

Our novel findings relative to the literature are accounted for by our use of high quality detailed industry data for firms from the Bureau of Labor Statistics (BLS) and the Bureau of the Census. Song et. al. (2019) use the SSA data which is very high quality matched employer-employee data in terms of tracking workers and firms but, as we discuss below, industry codes in the SSA data are known to be of inferior quality. Even with arguably inferior industry codes, follow up work by Bloom et. al. (2018) note that industry codes exhibit an especially high degree of missingness in their analysis. Acemoglu and Autor (2011) use household data (e.g., ACS and Decennial) with occupation codes and broad (one-digit) industry codes from self-reporting by households. Such codes are also known to be of inferior quality and one-digit is too coarse to capture the effects we detect.² In addition to taking advantage of high quality industry codes for firms from BLS and Census, we also take advantage of the Occupational Establishment Survey (OES) data that provides high quality occupation codes by high quality detailed industry codes.

¹ See page 22 of Song et. al. (2019) and page 1092 of Acemoglu and Autor (2011).

² See the discussion in Mellow and Sider (1983) and Dey et. al. (2010).

While we challenge the conventional wisdom that between industry effects are unimportant, in many respects we build on the findings of this literature. Acemoglu and Autor (2011) find especially large increases in employment and earnings for professional, technical and management occupations while occupations such as sales, production, operators and clerical workers experience declines. They suggest these patterns point towards a conceptual framework that emphasizes the distinction between tasks and skills. Consistent with their findings, we find that the occupations with the sharpest rise in earnings differentials that differ by industry are managerial, professional and technical with occupations like production and sales being left far behind. Our contribution is to integrate the changing role of occupations with the rise in between firm and in turn between industry inequality. Our findings show that the changing demand for tasks as evidenced in the changing occupational mix and occupational differentials is closely linked to between firm and between industry inequality.

Likewise, our findings build on the seminal findings of Song et. al. (2019). Beyond demonstrating that rising inequality is mostly accounted for by rising between firm inequality, they use an AKM decomposition of earnings to show that much of this increase is accounted for by increased sorting and segregation across firms. Our findings suggest that the increase in sorting and segregation is between industries as well as between firms. In addition, the increased sorting and segregation between firms and industries must be consistent with the dominant role of changing occupation mix and differentials between industries. Taken together, these findings provide substantial guidance about where to look for the ultimate sources of rising inequality.

For our analysis, we exploit the matched employer-employee data from the Longitudinal Employer Household Dynamics (LEHD) program at Census to revisit these issues along with other rich data sources at Census. These include access to the industry codes used by both BLS

and Census, as well as to the firm-level industry codes from SSA on a limited basis. We match the findings of the recent literature showing a dominant role for between firm effects in accounting for rising earnings inequality across workers. However, our findings depart from the SSA based analysis in finding the dominant role of between industry effects in accounting for the rising between firm inequality. In turn, the dominant role of industry effects enables us to integrate the OES data at the 4-digit NAICS level. This enables our investigation of the changing role of occupation effects within detailed industries.

The paper proceeds as follows. Section II describes the data infrastructure. Section III presents the main results of the paper showing the dominant role of between firm effects in accounting for rising overall earnings inequality, the dominant role of between industry effects in accounting for rising between firm inequality, and the dominant role of between occupation effects in accounting for between industry effects. Section IV presents an accounting decomposition. Our focus in this analysis is to decompose the contribution of industry and occupation effects into that accounted for by changing composition vs. changing earnings differentials by industry and occupation. As part of this analysis, we also investigate the contribution of industry effects in accounting for rising between firm earnings inequality with and without a rich set of firm controls. In turn, we also examine the contribution of occupation effects in accounting for rising inter-industry earnings differentials both without and with firm controls. Section V presents sensitivity analysis using alternative industry codes including the SSA codes (as Census obtains these codes from SSA as part of their processing). Section VI presents concluding remarks.

II. *Data Infrastructure*

The Longitudinal Employer Household Dynamics (LEHD) is a longitudinally linked employer-employee dataset created by the U.S. Census Bureau as part of the Local Employment Dynamics federal-state partnership. The data are derived from state-submitted Unemployment Insurance (UI) wage records and the 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 lists the quarterly earnings of every individual 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 LEHD is defined as the presence of an individual-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 various amounts of historical data upon joining the LEHD program, the length of the time series of LEHD data varies by state. The data used here are private sector data from the 20 states that have data available from 1996:Q1 through 2015:Q4.³

³ These 20 states are: CA, CO, CT, HI, ID, IL, KS, LA, MD, MN, MO, MT, NC, NJ, NM, OR, RI, TX, WA, and WY. These 20 states account for roughly 46 percent of national employment. The time series of employment from these 20 states closely tracks the national time series of total private sector employment published by BLS.

We use full quarter jobs in our analysis, where a full quarter job is defined as a contemporaneous employer-employee match that also exists in the previous quarter and in the following quarter. The underlying assumption is that individuals in full quarter jobs are working all 13 weeks of the quarter, which avoids the issue of not knowing the number the weeks worked during the quarter for individuals who start a job or end a job during that quarter. Restricting to full quarter jobs is similar in spirit to the full-time or full-year restriction used when analyzing inequality with household survey data. In the 1996:Q2 – 2015:Q3 time period, we have 4.9 billion jobs and 3.3 billion full quarter jobs in our 20 state LEHD data.

We impose several recodes to the LEHD earnings data. First, to minimize the effect of outliers and smooth the first two moments of the earnings time series, we topcode earnings at the 99.5th percentile of the state-year-quarter distribution. Second, all of our analysis uses the natural log of real earnings, where nominal values are converted to real using the 2015:Q3 CPI-U-RS deflator.

We merge the quarterly LEHD with the annual revenue-enhanced Longitudinal Business Database (RE-LBD). The RE-LBD is the Census Bureau's business register with information on a firm's real revenue per worker, which we refer to as labor productivity. In all our analysis, we use the enterprise as the definition of the firm, where an enterprise is defined as an aggregation of Employer Identification Numbers (EINs). The merged LEHD – RE-LBD data contains annual observations for the years 1997 through 2013.⁴

Using annual observations in our analysis requires us to choose an earnings measure from the quarterly LEHD. We use Q1 full-quarter earnings in each year. Our key findings in Table 1 are not sensitive to whether we use Q1, Q2, Q3, or Q4 full quarter earnings. Furthermore, many

⁴ The source data for revenue in the RE-LBD are sometimes missing, and we use a propensity score weight to correct for this issue.

of our key findings are not sensitive to whether we sum the LEHD quarterly earnings into an annual measure of earnings with a minimum earnings threshold. Annual earnings are used by Song et.al (2019) using SSA data, as well as by Abowd et.al (2018) using LEHD data. The key results in Table 1 change dramatically when no minimum earnings threshold is applied to annual earnings data, most likely due to a decline in short-duration jobs and thus a compositional change in the lower part of the earnings distribution during the 1997 – 2013 time period (see Hyatt and Spletzer, 2017). Song et. al. (2019) also impose a minimum earnings threshold.⁵

Industry codes play a fundamental role in our analysis. Our basic results use establishment-level industry codes from the BLS QCEW program, aggregated to the enterprise level. Aggregation from establishment level data is done using maximum employment (for example, if an enterprise has $N \geq 1$ establishments with $N \geq M \geq 1$ industry codes, the industry code with the maximum employment is chosen for the aggregation). In some of our sensitivity analysis, we use industry codes from the Census’ RE-LBD and from the Social Security Administration – these alternative industry codes are described and analyzed in section V.

Almost all of our analysis uses the variance as the measure of earnings inequality.⁶ This is done because of the simple formula where the variance of individual earnings can be decomposed into variance within firms and variance between firms:

$$\text{var}(W_{if}) = \text{var}(W_{if} - W_f) + \text{var}(W_f) \quad (1a)$$

where “i” refers to the individual and “f” refers to the firm. Furthermore, letting “k” refer to industry, we can further write this variance formula as:

⁵ In unreported results, we find that our results are very similar if we use exactly the same type of minimum earnings threshold as Song et. al. (2019).

⁶ We have also calculated the {10th, 50th, 90th} percentiles from the LEHD full quarter earnings and find they are similar to the published {10th, 50th, 90th} percentiles of weekly earnings of full-time workers from the Current Population Survey (CPS) (see Appendix Figure A.1).

$$\text{var}(W_{ifk}) = \text{var}(W_{ifk} - W_{fk}) + \text{var}(W_{fk} - W_k) + \text{var}(W_k). \quad (1b)$$

Calculating this variance decomposition in each year, and letting Δ denote changes across time, we have

$$\Delta \text{var}(W_{ifk}) = \Delta \text{var}(W_{ifk} - W_{fk}) + \Delta \text{var}(W_{fk} - W_k) + \Delta \text{var}(W_k). \quad (1c)$$

The increase in the variance of individual level wages can be decomposed into a change within firms (the first term on the right hand side of equation 1c), the change between firms within industries (the second term), and the change between industries (the third term).

The variance components in equation (1c) can be estimated multiple ways. They can be computed directly from the LEHD individual-level data – this is what we do in Table 1. The firm-level and industry-level terms on the right hand side of equation (1c) can also be computed from firm-level data with appropriate employment weights. The industry-level term can also be computed from firm-level employment-weighted regression analysis with industry fixed effects – this is what we do in Table 2.

III. *Main Results: The Dominant Role of Industry and Occupation Effects*

Table 1 presents the main results of the paper. Reported in Table 1 are a series of variance decompositions of the changes in earnings inequality. The top panel depicts the variance of (log) earnings per worker within and between firms (using the enterprise definition of firms described above) and within and between industries in 1997. The estimates in the top panel of Table 1 coincide exactly with the terms in equations (1a) and (1b) above. The middle panel presents an analogous variance decomposition for 2013, and the third panel is merely a first difference of the estimates in the top two panels. Between 1997 and 2013, overall earnings inequality increased by about 17.2 log points (the third panel of Table 1). About 74% of this

increase is accounted for by the rapidly rising between firm inequality. This finding is broadly similar to the Song et. al. (2019) finding that about 70% of the increase in earnings inequality from 1981 to 2013 is due to between firm effects (where firms are defined by EINs).^{7,8}

Most of the increase in between firm dispersion is accounted for by an increase in between industry dispersion (third panel of Table 1). Using 4-digit NAICS effects, the fraction of the rise in between firm earnings dispersion accounted for by between industry effects is about 75%. Even 2-digit effects account for 38% of the rise in between firm inequality, 4-digit effects account for about 75% of the increase and 6-digit effects account for 82% of the increase. Using the OES occupation shares by detailed 4-digit industry, the fourth panel of Table 1 shows that 92.1% of the rising between industry dispersion is accounted for by occupation effects.⁹

In short, the striking finding from the recent literature is that most of the rise in individual earnings inequality is due to rising dispersion between firms. We confirm this using the LEHD data. Our contribution is to show that a substantial amount of this increase in between firm dispersion is due to rising dispersion between industries. In turn the latter is almost completely accounted for by occupation effects. These findings imply that explanations of rising inequality need to account for jointly the connection between firm, industry and occupation effects. In the

⁷ Barth et al. (2016) do not emphasize it but in using the Longitudinal Business Database they find that about 50% of the increase in between establishment earnings is due to between 4-digit SIC effects for the period 1977 to 2007. Several differences stand out relative to our study: their earlier time period, their use of establishment-level as opposed to firm-level data, and importantly their use of SIC codes. The switch from SIC to NAICS yielded much more detail in the information technology industries outside of manufacturing.

⁸ For our sample period, we find that the fraction of the rise in earnings dispersion accounted for by employer effects increases with using a narrower definition of business. For example, we find that if we use the State UI Account number (SEIN) that over 90% of the increase in earnings dispersion is accounted for by between SEIN effects (see Appendix Figure A.2 for details).

⁹ As we describe later in section IV.B, the occupation results require restricting to a slightly narrower set of industries covered by the OES and covers only the 2002-13 period. However, the 19 industries in the LEHD not published by the OES are very small, and the overall 2002-2013 increase in variance growth between 4-digit industries is exactly the same for the larger (304) and slightly smaller (285) sample of industries. The variance accounted for by occupation effects is from an employment-weighted regression of the industry-level earnings differentials on shares of employment in each of the 22 occupation groups.

next section, we provide further details to provide guidance for such explanations. First, we explore whether the rising contribution of industry effects is derived from changes in the mix of employment across industries or changes in inter-industry earnings differentials. We find that, with and without additional firm controls, most of the rising contribution of industry effects is due to rising inter-industry earnings differentials. In turn, we explore whether the contribution of occupation effects in accounting for the rising inter-industry earnings differentials is due to changes in the occupation mix across industries or changes in occupational earnings differentials. For this analysis, we further explore the patterns that emerge for specific industries and occupations.

IV. *Changing Quantities or Prices?*

A. *Industry Effects*

The dominant role of industry effects could be from either changing composition of employment across industries with different average earnings or changing earnings differentials. To evaluate this, we use the variance accounting methodology of Juhn, Murphy and Pierce (1993) (hereafter JMP). We don't pursue the full distribution accounting insights from this approach but focus on the decomposition of variance.¹⁰ This approach enables us to not only gain insights into the "price" vs. "quantity" effects but also to investigate whether controlling for firm-level observables accounts for the observed industry effects.¹¹

To be specific, we estimate the following two specifications:

¹⁰ There are some limitations of the JMP decomposition methodology as highlighted by DiNardo et. al. (1996) and Fortin et. al. (2010). These limitations primarily apply to the full distribution accounting (e.g., decomposing the 90-50 vs. the 50-10) which we do not pursue.

¹¹ The ordering of the decomposition potentially matters. We have used the standard approach first accounting for changing quantities, then prices and finally the residual. We have found that results are robust to changes in the ordering of the quantity vs. price effects.

$$Y_{ft} = X'_{ft}\bar{\beta} + Ind'_{ft}\bar{\gamma} + \varepsilon_{ft} \quad (2a)$$

$$Y_{ft} = X'_{ft}\beta_t + Ind'_{ft}\gamma_t + \tilde{\varepsilon}_{ft} \quad (2b)$$

Specification (2a) is a pooled specification with time invariant coefficients with the dependent variable earnings per worker at the firm-level in each year, X is a vector of firm level controls, Ind is a vector of industry effects (using 4-digit industries). Specification (2b) permits the effects of the observables and industry effects to vary over time. All specifications are estimated on an employment-weighted basis. Specification (2b) can be rewritten as follows:

$$Y_{ft} = X'_{ft}\bar{\beta} + Ind'_{ft}\bar{\gamma} + X'_{ft}(\beta_t - \bar{\beta}) + Ind'_{ft}(\gamma_t - \bar{\gamma}) + \tilde{\varepsilon}_{ft} \quad (2c)$$

Following JMP, Davis and Haltiwanger (1991) and Dunne et. al. (2004), the changes in dispersion (either the variance or other moments) can be decomposed into quantity (X and Ind) effects for average prices ($\bar{\beta}, \bar{\gamma}$), price effects (β_t and γ_t) and the residual. The estimation and decomposition is on an employment-weighted basis to be consistent with the variance decompositions reported in Table 1.

The first column of Table 2 shows results for a regression with just 304 industry fixed effects as explanatory variables. The R-squared for the pooled regression is .5429, implying that knowing the 4-digit industry explains 54% of between firm earnings dispersion. Not reported in Table 2 but informative are the R-squareds for the year specific regressions (corresponding to equation 2b) – these are .5363 for 1997 and .5813 for 2013. When multiplied by the corresponding between firm variances as reported in Table 1, these year-specific R-squareds show the amount of between-firm variance that is between industries (.5363 * .485 = .260 and .5813 * .613 = .356).

The lower panels of Table 2 show the results of the JMP decomposition. First, looking at the first column of Table 2, we readily observe that this regression based variance decomposition

replicates the finding in Table 1 that industry effects account for about 75% of the rise in between firm dispersion (as measured by the variance). The additional insight from this analysis is that only 10% of the 75% is due to changing composition of employment across industries while 65% is due to changing inter-industry earnings differentials.

To shed further light on the contribution of industry-effects, we also consider a range of firm-level controls. Specifically, we include firm-level LN real revenue per worker (labor productivity), firm age and firm size, and worker mix controls – the share of workers at the firm by worker age (seven age groups), gender, and education (four education groups).¹² We interpret these as capturing several possible factors underlying the change in firm-level earnings dispersion. The inclusion of labor productivity captures effects associated with rent-sharing but might also reflect unobserved worker quality effects. Firm size and age are correlated with heterogeneity in firm performance. Worker characteristics such as age, gender and education capture differences in the choice of worker mix across firms.

Results from this supplementary analysis are reported in column (2). We find that earnings per worker are higher at firms that have higher labor productivity, are larger, older, have a higher share of workers in the 25-34 age range, a higher share of males, and more educated workers. Of more direct interest are the results of the variance accounting decompositions reported in the lower panels. Even with all controls, industry effects account for 48% of the rise in between firm inequality. Most of this (41% of the 48%) is due to rising residual inter-industry earnings differentials rather than changing employment composition across industries.

¹² The LEHD education is directly measured for individuals that have appeared in the 2000 Decennial and ACS. For other workers it is imputed. Our main findings are robust to exclusion of this variable. We include it to highlight that the dominant role of industry is robust to inclusion of this control.

Taking stock, most of the rise in the contribution of industry effects in accounting for rising between firm inequality is due to a rise in inter-industry earnings differentials. Even with extensive firm-level controls such as firm performance and firm-worker type composition, the rise in inter-industry earnings differentials still accounts for most of the increase in the contribution of between industry effects.¹³

B. *Occupation Effects*

To investigate the role of occupations further, we estimate an analogous set of regression specifications to (2a) and (2b) but now with the dependent variable the industry-level earnings differential and the explanatory variables the employment shares by occupation and detailed industry from the OES. Using the same type of JMP style decomposition, we decompose the rise in dispersion in inter-industry earnings differentials into quantity and price effects. That is, the contribution of changing occupational distribution across industries and changing occupation differentials for a given difference in occupational distributions across industries. For these exercises, we consider two alternative dependent variables. First, the raw inter-industry earnings differentials without firm-level controls. Second, the residual inter-industry earnings differentials with all firm-level controls.

The occupation data by 4-digit industry are downloaded from the BLS website. The Occupational Employment Statistics (OES) program at BLS publishes employment for 22 occupations for approximately 290 NAICS industries in each year 2002 forward. The 22 occupations are listed in Table 3. After excluding industries 9991 and 9993, there are 285 4-digit

¹³ Appendix Figure A.3 depicts the 3-digit NAICS industries with the largest increases (panel A) and decreases (panel B) in inter-industry residual earnings differentials (relative to the mean) after controlling for all firm-level controls. Also, reported are the changes in inter-industry earnings differentials for these same industries without controls. The rank ordering and even magnitudes of the changes in inter-industry differentials are broadly similar with and without controls (Appendix Figure A.4). There is less dispersion in the change in inter-industry earnings differentials with controls and the slope in Appendix Figure A.4 is less than one – both of these findings are consistent with firm-level controls absorbing some but not most of the rising inter-industry earnings differentials.

industries published in the OES each year 2002 to 2013. The 19 industries in the LEHD data but not in the OES data are mostly in agriculture. The occupation variables on the right hand side of the industry-level earnings regressions are the percentage of industry employment in the 22 occupations.¹⁴

Results from this exercise are depicted in Table 3. The basic results mimic those of Table 1.¹⁵ Using the raw differentials, occupation effects account for 92% of the rise in dispersion from inter-industry earnings differentials. However, beyond Table 1, the accounting decomposition indicates that 66% of this is due to changing occupation distributions across industries and 27% due to rising occupation differentials. Interestingly, the basic findings in Table 3 using the raw differentials can be replicated with OES public-domain data (see appendix B). This highlights one of the reasons that the dominant role of industry effects is important. Analysis of potential sources can be done with public domain data. Also, it highlights that our findings are not being driven by the use of the LEHD earnings measures.

With the residual differentials, occupation effects account for 94% of the rise in between industry dispersion. 26% of this is due to changing occupation distributions across industries and 68% due to rising occupation differentials. In short, we account for virtually all of the rise in inter-industry earnings differentials with occupation effects (with or without controls).

Table 3 also reports the changing occupation differentials for individual occupations. We focus on the last column since this includes the firm controls. Since the latter controls for the age, gender, and education mix of the workforce, this is closer to the Acemoglu and Autor

¹⁴ The OES publishes employment for the 4-digit industry total, as well as for the 22 occupations in the industry. We create a 23rd occupation category titled “other” which is total employment minus the sum of the 22 published occupations.

¹⁵ The supplemental decomposition of the contribution of occupation effects into “price” and “quantity” effects is pushing the data hard given the sample size in Table 3. The F-tests in Table 3 provide reassurance that this supplemental decomposition yields meaningful insights.

(2011) findings that report changing occupation differentials with a wide range of similar controls at the person level. We find sharp increases especially in management but also substantial increases for business and financial, sciences, legal, and healthcare occupations. In contrast, occupations such as protective services, sales, and construction experience declines. These patterns are broadly consistent with those reported by Acemoglu and Autor (2011). Our contribution is that these changes in occupation differentials reflect changes in occupation differentials across industries.

The especially large increases in the occupation differential for management is especially striking. To explore this further, Figure 1 shows the scatter plot of the management share of employment by 4-digit industry (x-axis) vs. the inter-industry earnings differentials by industry. The scatterplot includes both 2002 (black) and 2013 (red) data. We find that it is in industries with especially large management shares that there is the sharpest rise in inter-industry earnings differentials between 2002 and 2013. It is this variation that is driving the findings in Table 3 for management.¹⁶

Further insights on the rising management differential emerge by considering a modified regression specification that includes interactions between the management occupation employment shares and industry dummies.¹⁷ We find (see Figure 2) that the management differential rises in almost all industries with the exception of Retail Trade. The industries with especially large increases in management differentials include Information, FIRE, Professional and Technical Services and Education.

¹⁶ The patterns in Figure 1 are robust to using the inter-industry earnings differentials without firm controls (Appendix Figure A.5).

¹⁷ For this specification we use 15 sectors as depicted in Figure 2 for the interactions.

V. *Sensitivity Analysis on the Quality of Industry Codes*

As discussed above, both BLS and Census have strong incentives and extensive statistical programs to assign detailed and accurate industry codes at the establishment-level. For BLS, this includes the QCEW program along with the Annual Refiling Survey. 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)).

Here we explore differences in the information content of industry codes from BLS vs. Census vs. SSA for 2013. We are limited in our analysis of using SSA codes because prior to 2002 Census did not store the SSA industry codes once replaced by either Economic Census or BLS codes (and after an Economic Census all codes have been replaced for businesses in that Census). In 2013, we have BLS, Census and SSA codes for all firms that are 14 years old or less. For this sample, we compute the fraction of the cross sectional between firm dispersion in earnings accounted for by industry effects using Census, BLS and SSA industry codes (at the 4-digit NAICS level). Using Census and BLS codes, we find that about 45% of the between firm variance is accounted for by between industry variation. Using the SSA codes, between industry variation accounts for only 25% of the between firm variance. Even though this is a limited exercise, it highlights that the information content of SSA codes in terms of accounting for variation in earnings across firms is much smaller than that for BLS and Census.

It still might seem surprising that Song et. al. (2019) attribute only a negligible fraction of the increase in between firm variance in earnings over time to between industry variation. However, Bloom et. al. (2018) indicates that the same SSA micro data used in Song et. al. (2019) has missing industry codes for all new firms post 2002. Their Table 2 shows that this yields that EINs with missing industry codes increased from accounting for only 4% of total employment in the 1980-86 period to 24% in 2007-13 in their micro data. Our inference from our analysis is that even without this substantial missingness problem, the use of SSA codes yields a misleading characterization of the role of between industry variation in accounting for between firm variation.

VI. *Concluding Remarks*

Our findings integrate two distinct approaches and sets of results regarding rising earnings inequality. Acemoglu and Autor (2011), both in original research and in summarizing the literature using household surveys, highlights the increasing role of the changing occupational mix and returns in accounting for rising earnings inequality post 1980. Alternatively, Song et. al. (2019) highlights the dominant role of rising between firm inequality in accounting for rising earnings inequality over this same period.

Our findings highlight that these results are very closely related in a perhaps somewhat surprising manner. Specifically, we find that the rise in between firm inequality is dominated by a rise in between industry inequality. We find the latter is dominated by rising inter-industry earnings differentials with a more modest role for the changing distribution of employment across industries. In turn, we find that virtually all of the rising inter-industry earnings differentials is accounted for by occupation effects. This includes both changing mix of

occupations within and across industries and rising occupation differentials. We find, after controlling for the mix of workers in terms of age, gender, and education (as well as other controls), that the occupations with especially sharp increases in earnings differentials that differ across industries are managerial, business and financial, sciences, legal, and healthcare, with occupations like protective services, sales, and construction being left behind. The increase in the management occupation differential across industries is especially striking.

The somewhat surprising nature of our findings is that these results are related via structural changes at the industry level. Both of the prior literatures missed this finding essentially because of relatively limited industry codes in household data and in the SSA firm-level data. By using high quality industry codes from Census and BLS and in turn high quality industry by occupation data from BLS we have uncovered this connection.

Our findings suggest that to be able to understand the changing structure of work across firms and occupations that one needs to understand the changing interaction between industries and occupations. Our findings suggest that the framework developed by Acemoglu and Autor (2011) to understand the changing occupation effects should include an important industry-specific component. The framework they outline has production activities requiring tasks and different skills having comparative advantage for alternative tasks. This task based framework lends itself nicely to considering changes in the structure of production induced by technology, globalization and other factors. Put differently, this task based framework helps motivate why changes in inequality associated with the changing role of occupations is linked to the changing role of firms in inequality. Driving forces such as technology and globalization that induce changing demand for occupations impact firms differentially. In principle, this differential impact could be between firms or between industries. Our findings suggest there is a dominant

role for the between industry impact. An important direction for future research is to explore the changing role of occupations across industries.

References

- Abowd, John M., Bryce E. Stephens, Lars Vilhuber, Fredrik Andersson, Kevin L. McKinney, Marc Roemer, and Simon Woodcock. 2009. "The LEHD Infrastructure Files and the Creation of the Quarterly Workforce Indicators," in *Producer Dynamics*, Timothy Dunne, J. Bradford Jensen, and Mark J. Roberts, eds., Chicago: University of Chicago Press, pp. 149-230.
- 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*, Vol. 36, No. 51, pp. S183-S300.
- Acemoglu, Daron and David H. Autor. 2011. "Skills, Tasks and Technologies: Implications for Employment and Earnings." in Orley Ashenfelter and David Card, eds., *Handbook of Labor Economics*, Volume 4, Amsterdam: Elsevier-North Holland, pp. 1043-1171
- 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 (2, pt. 2): pp. S67-S97.
- Bloom, Nicholas, Fatih Guvenen, Benjamin S. Smith, Jae Song, and Till von Wachter. 2018. "The Disappearing Large-Firm Premium." *American Economic Review Papers and Proceedings*, 108: pp. 317-322.
- Davis, Steve J., and John Haltiwanger. 1991. "Wage dispersion between and within U.S. manufacturing plants, 1963–1986." *Brookings Papers on Economic Activity: Microeconomics*, pp. 115–200.
- Dey, Matthew, Susan Houseman, and Anne Polivka. 2010. "What Do We Know About Contracting Out in the United States? Evidence from Household and Establishment Surveys" in *Labor in the New Economy*, Katharine G. Abraham, James R. Spletzer, and Michael Harper, eds., Chicago: University of Chicago Press, pp. 267-304.
- DiNardo, John, Nicole M. Fortin, and Thomas Lemieux. 1996. "Labor Market Institutions and the Distribution of Wages, 1973-1992: A Semiparametric Approach", *Econometrica* 64: 1001-1044.
- 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*, pp. 397-429.
- Fortin, Nicole, Thomas Lemieux and Sergio Firpo. 2010. "Decomposition Methods in Economics" NBER Working Paper No. 16045, June.

Hyatt, Henry R. and James R. Spletzer. 2017. "The Recent Decline in Single Quarter Jobs." *Labour Economics*, Vol. 46, pp. 166-176.

Juhn, Chinhui, Kevin M. Murphy, and Brooks Pierce. 1993. "Wage inequality and the rise in returns to skill." *Journal of Political Economy* 101 (June): 410–42.

Mellow, Wesley, and Hal Sider. 1983. "Accuracy of response in labor market surveys: Evidence and implications." *Journal of Labor Economics* 1, no. 4:331–44.

Song, Jae, David J. Price, Fatih Guvenen, Nicholas Bloom, and Till von Wachter. 2019. "Firming Up Inequality." *Quarterly Journal of Economics*, Vol. 134, No. 1, pp. 1–50.

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.

Table 1: Variance Decomposition, log real quarterly earnings

1.094	Variance, 1997
.609	Variance within firms
.485	Variance between firms
.225	Variance between firms, within 4-digit industries
.260	Variance between 4-digit industries
1.266	Variance, 2013
.653	Variance within firms
.613	Variance between firms
.257	Variance between firms, within 4-digit industries
.356	Variance between 4-digit industries
.172	Variance growth, 1997-2013
.044	Variance growth within firms
.128	Variance growth between firms 74.4% of total variance growth is between firms
.032	Variance growth between firms, within 4-digit industries
.096	Variance growth between 4-digit industries, 1997-2013 75.0% of variance growth between firms is between industries 55.8% of total variance growth is between industries
.089	Variance growth between 4-digit industries, 2002-2013
.082	Variance growth between 4-digit industries explained by occupation 92.1% of variance growth between industries is explained by occupation

Notes: [a] LEHD full quarter Q1 earnings from 20 states.

[b] Number of weighted individuals is 34,070,000 in 1997 and is 44,030,000 in 2013.

[c] Number of weighted firms is 1,807,000 in 1997 and is 2,033,000 in 2013.

[d] Number of 4-digit industries is 304 in the top 3 panels of this table, and is 285 in the bottom panel of this table (see text for further explanation)..

[e] Mean LN real quarterly earnings is 8.983 in 2002 and is 8.960 in 2013.

Table 2: Firm level regressions
Dependent variable is log real quarterly earnings

	(1)	(2)
Labor Productivity		.1015
LN(1 + Firm Age)		.0365
LN(1 + Firm Size)		.0353
% Workers in Firm:		
Age < 16		-.8916
Age 16-24		-1.2070
Age 25-34		.0545
Age 45-54		-.0866
Age 55-64		-.2662
Age > 65		-1.1110
Female		-.5674
Education < 12		.0082
Education 13-15		.4344
Education 16+		1.1820
Industry Fixed Effects	Yes	Yes
R-Squared	.5429	.7141
JMP contributions:		
Changing X	10.1%	-3.4%
Changing β	65.1%	76.6%
Residual	24.9%	26.8%
Contribution of industry effects to variance of earnings growth		
Contribution of X only	10.0%	7.3%
Marginal contribution of β	65.1%	40.6%
Overall contribution	75.1%	47.8%

Notes: [a] LEHD full quarter Q1 earnings from 20 states, aggregated to the firm level, all years 1997-2013.

Sample size is 25,670,000 firm-year observations.

[b] Regression weighted by RE-LBD weight.

[c] There are 304 industries.

[d] The calculation of “JMP contributions” and “Contribution of industry effects to variance of earnings growth” is explained in the text.

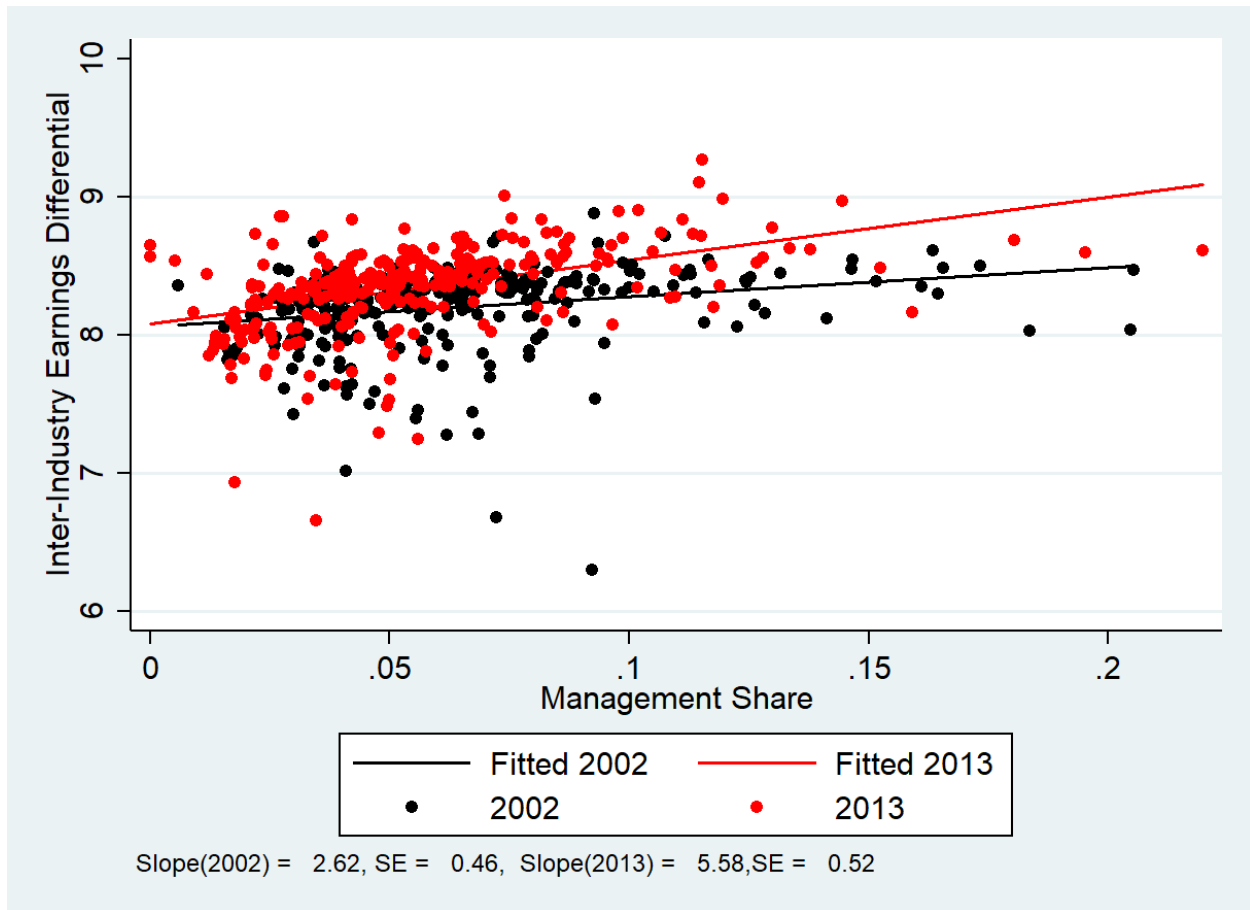
[e] All standard errors in column (2) are less than .003, and all t-statistics in column (2) are greater than 8.

Table 3: Industry level regressions
Dependent variable is log real quarterly earnings

Percent of industry employment in occupation:	All years Mean	Industry fixed effects from regression with no controls		Industry fixed effects from regression with all firm controls	
		All years β	$\beta(2013)-$ $\beta(2002)$	All years β	$\beta(2013)-$ $\beta(2002)$
Management	4.8%	12.677	1.226	10.140	1.332
Business and Financial	4.0%	9.490	-0.142	8.506	0.447
Computer and Math	2.7%	10.182	-0.199	8.466	0.163
Architecture and Engineering	2.3%	9.792	0.202	8.389	0.288
Life Physical Social Science	0.8%	10.593	-0.600	8.702	0.331
Community and Social Services	0.9%	7.935	0.350	7.644	0.216
Legal	0.6%	9.755	-0.284	8.770	0.300
Education	1.6%	8.264	-0.051	7.813	0.191
Arts, Design, Entertainment	1.4%	7.530	0.011	6.527	-0.007
Healthcare	6.1%	9.522	0.184	8.421	0.318
Healthcare Support	3.2%	8.094	-0.253	8.275	0.013
Protective Services	1.0%	8.478	-0.112	7.885	-0.079
Food Preparation and Serving	9.0%	7.891	0.070	7.893	0.102
Building and Grounds	3.0%	8.137	0.023	7.806	0.030
Personal Care and Services	2.5%	7.379	-0.101	7.453	0.077
Sales	12.8%	8.197	-0.150	7.804	-0.120
Office and Admin	16.9%	8.985	0.126	8.360	0.099
Farming, Fishing, Forestry	0.2%	8.124	0.278	7.754	0.243
Construction	4.9%	8.864	-0.125	8.272	-0.074
Installation and Maintenance	4.3%	9.631	-0.077	8.521	0.095
Production	9.1%	9.104	0.016	8.344	0.021
Transportation	7.5%	8.860	-0.031	8.016	0.025
Other	0.4%	8.365	-6.670	6.825	-3.357
R-Squared		.801		.615	
F test for all coefficients equal		620.0	0.92	246.6	2.09
p value of F test		0.000	0.567	0.000	0.003
JMP contributions:					
Changing X		65.7%		25.8%	
Changing β		26.7%		67.8%	
Residual		7.6%		6.4%	

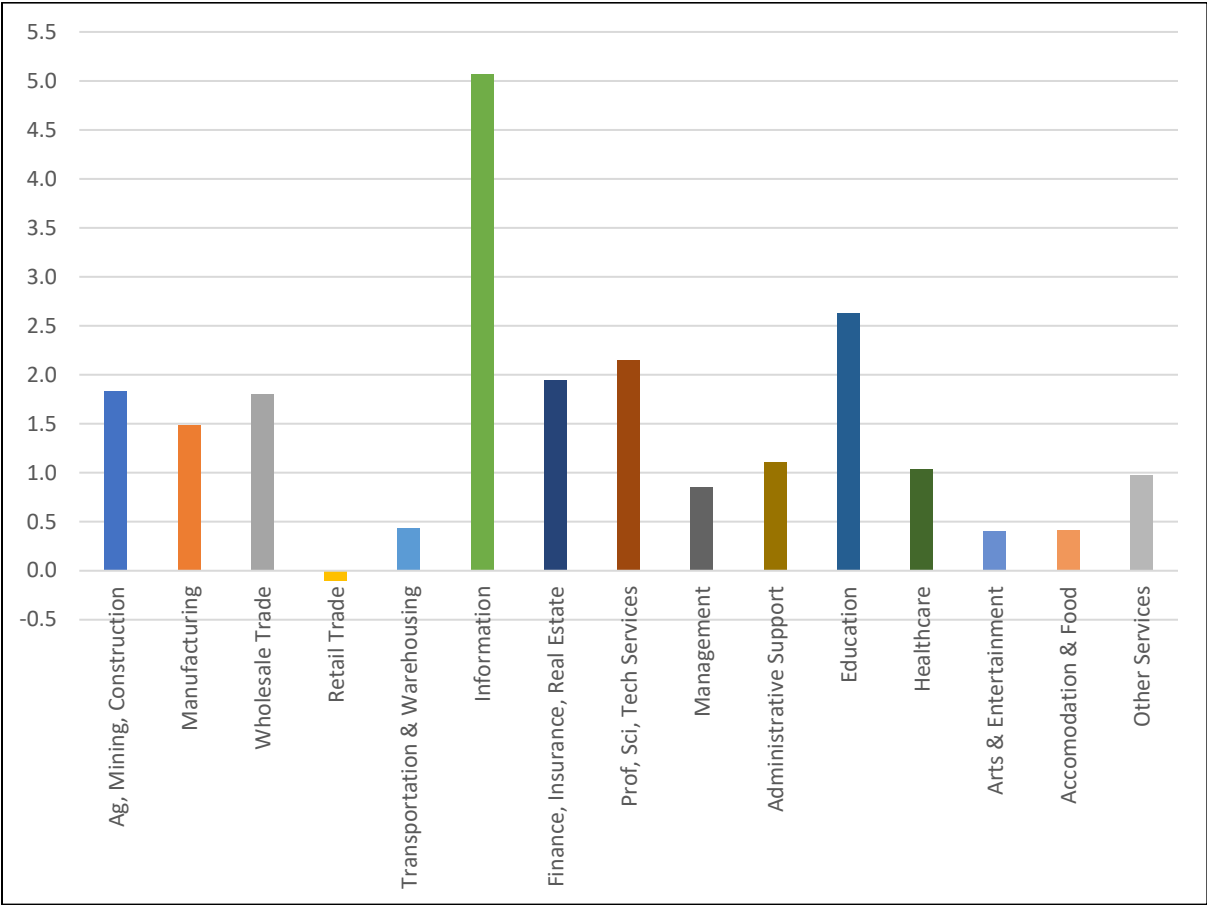
Notes: [a] LEHD full quarter Q1 earnings from 20 states, aggregated to the 4-digit industry level, merged to OES public use data, 2002-2013. Sample size is 3420 industry-year observations (285 industries x 12 years).
[b] Regression weighted by RE-LBD full-quarter industry employment.
[c] See footnote 14 for the definition of Other.
[d] Regression coefficients “All years β ” are from a specification with no intercept. The R-squared is from a regression with an intercept.
[e] $\beta(2013) - \beta(2002)$ is the difference in coefficients from 2002 and 2013 year-specific regressions.
[f] The calculation of “JMP contributions” is explained in the text.

Figure 1 The Relationship between Management Share and Inter-industry Earnings Differentials



Note: Reported inter-industry earnings differentials are those after using firm controls. The analogous figure without firm controls is in Figure A.5. Reported slopes using employment-weighted regressions.

Figure 2 Changing Management Earnings Differential by Industry

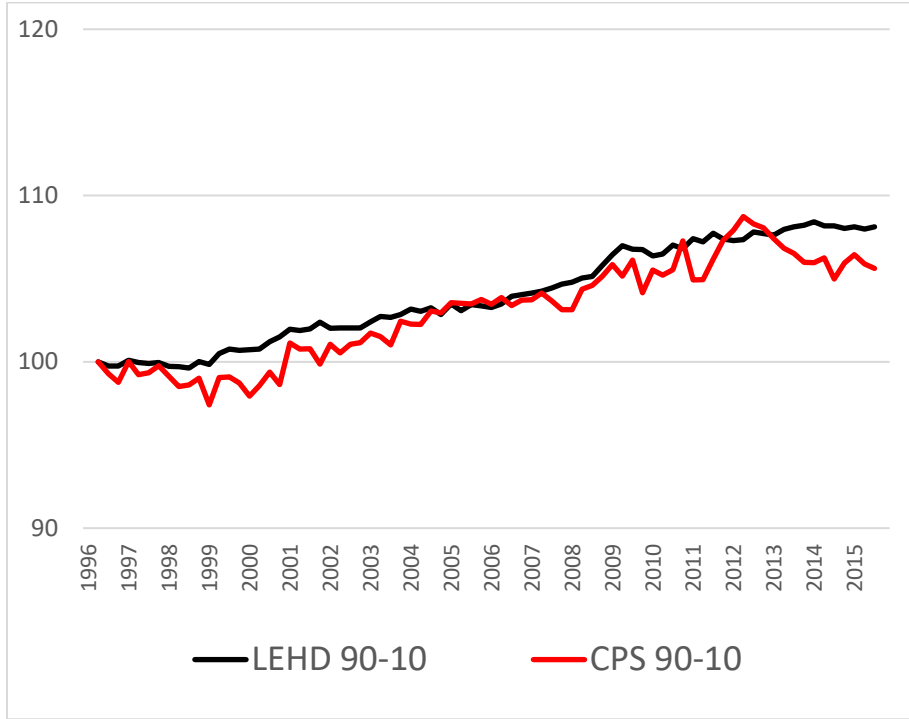


Note: Reported results are based on extended specification permitting occupational differentials to vary by industry. See text for details.

Appendix A: Tables and Figures

Figure A.1 Comparison of CPS and LEHD Changes in Earnings Inequality

A. 90-10 Differential



B. 90-50 and 50-10 Differential

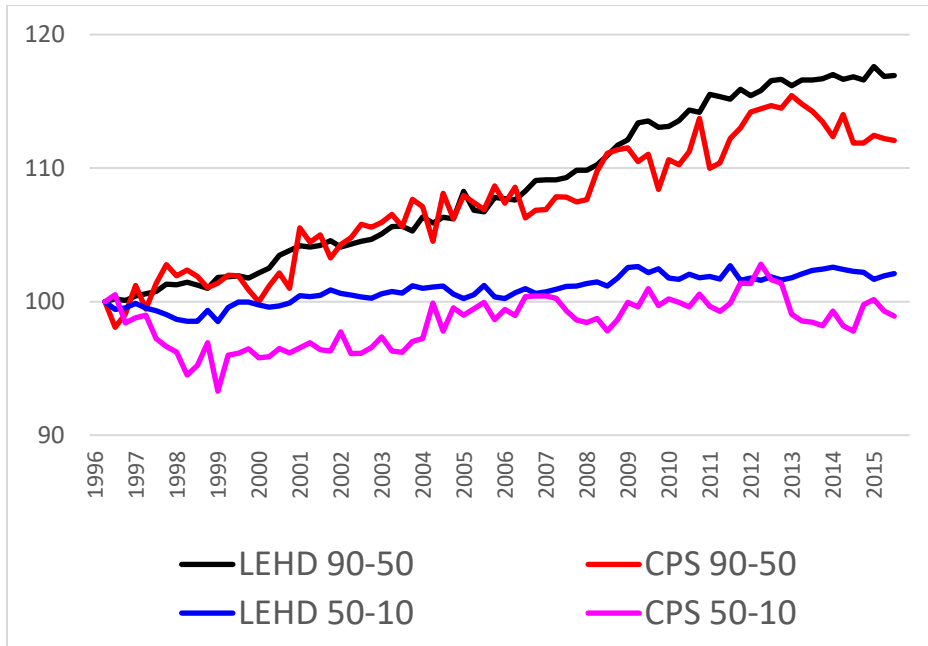
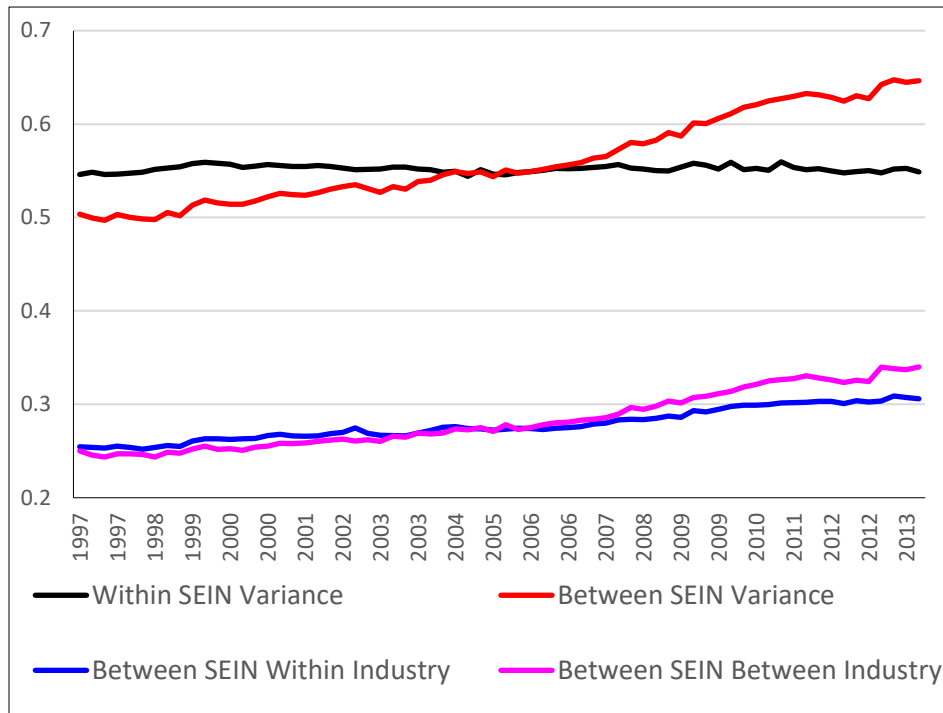


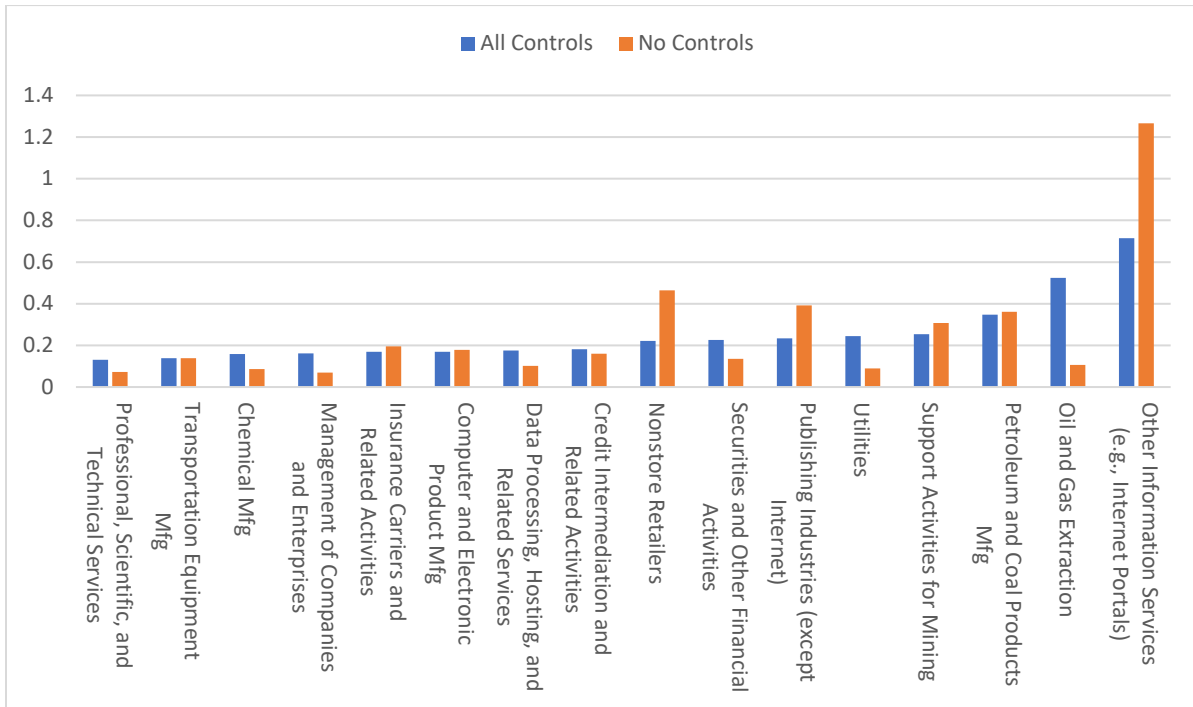
Figure A.2 Robustness to Alternative Definition of Businesses



Note: The SEIN is the state employer identification number from the LEHD data. It is more disaggregated than Census firm IDs but less disaggregated than establishment identifiers.

Figure A.3 3-digit NAICS Industries with largest changes in inter-industry earnings differentials

A. Industries with Largest Increases in Industry Earnings Differentials, 3-digit NAICS



B. Industries with Largest Declines in Industry Earnings Differentials, 3-digit NAICS

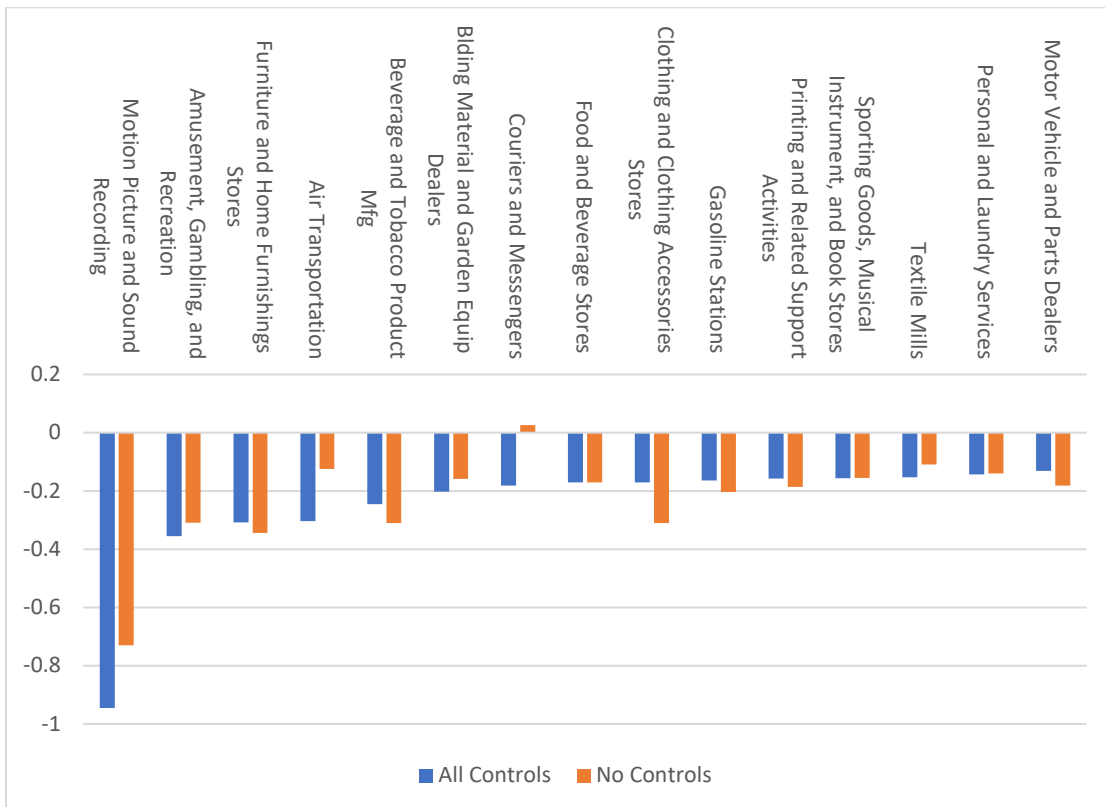
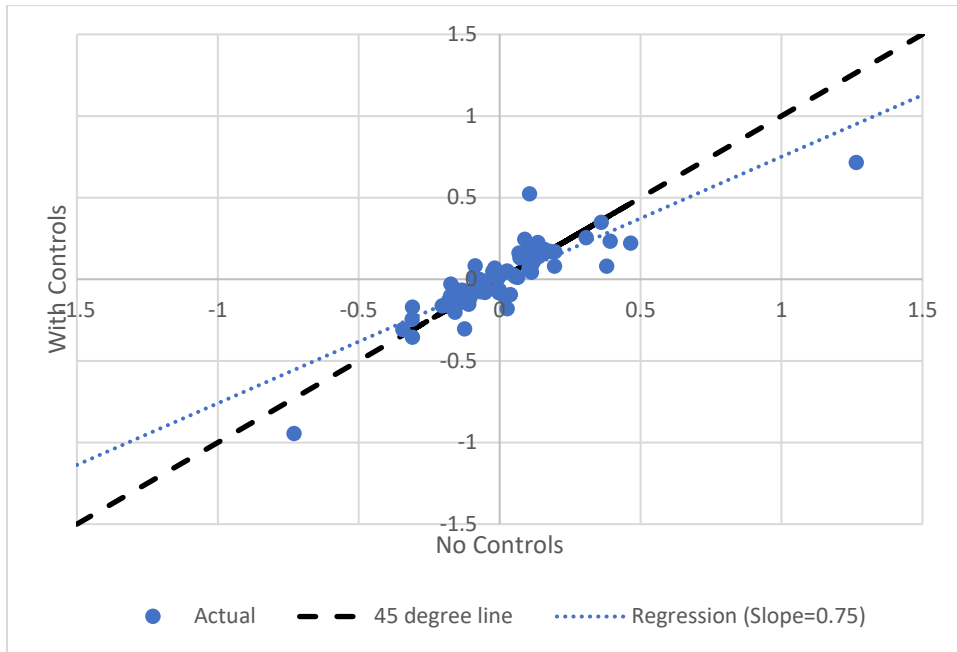
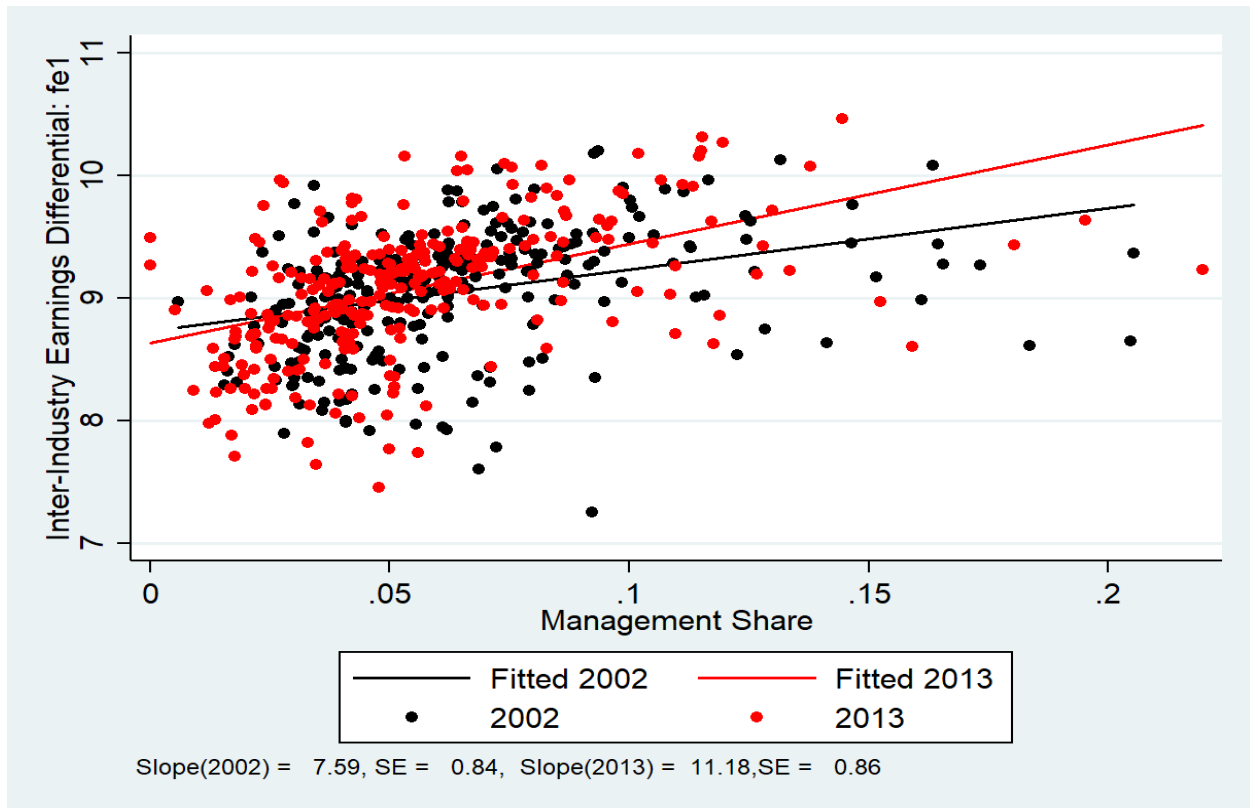


Figure A.4: Changes in Industry Earnings Differentials with and without controls (3-digit)



Note: Std Dev of Changes in Earnings Differentials (No Controls) = 0.23; Std Dev of Changes in Earnings Differentials (All Controls) = 0.20.

Figure A.5 The Relationship between Management Share and Inter-industry Earnings Differentials (no firm controls)



Appendix B. Replicating Table 3 results using OES public-domain data

Using the OES public domain data (which includes earnings by industry by year), we can replicate many of the findings in Table 3. First, we note that the unweighted correlation of OES industry earnings with our unadjusted LEHD earnings measure is .773. Second, using the OES earnings measure rather than the LEHD earnings measure in the Table 3 regression yields an R-squared of .866 (compared to 0.801 in Table 3) and very similar coefficients on the occupation measures. Furthermore, we get an R-squared of .858 using OES earnings as the dependent variable and OES employment (rather than LEHD FQ jobs) as the weight in the regression. In this latter regression that uses just OES data, the accounting decomposition indicates that 97% of the increase in inter-industry earnings differentials is due to changing occupation distributions across industries and 14% due to rising occupation differentials, with a negative residual component of -12 percent.