Firming Up Inequality*

Jae Song†
David J. Price‡
Fatih Guvenen§
Nicholas Bloom¶
Till von Wachterk

Abstract

We use a massive, new, matched employer-employee database for the United States to analyze the contribution of firms to the rise in earnings inequality from 1978 to 2013. We find that one-third of the rise in the variance of (log) earnings occurred within firms, whereas two-thirds of the rise occurred between firms. We show that the firm-related rise in the variance can be decomposed into two roughly equally important forces—a rise in the assortative matching of high-wage workers to high-wage firms and a rise in segregation of similar workers between firms. In contrast, we do not find a rise in the variance of firm-specific pay once we control for worker composition. The rise in the firm-related inequality due to worker sorting and segregation accounted for a particularly large share of the total increase in inequality in smaller and medium firms (explaining 84% for firms with fewer than 10,000 employees). In contrast, in the very largest firms with 10,000+ employees, almost half of the increase in the variance of earnings took place within firms, driven by both declines in earnings for employees below the median and a substantial rise in earnings for the 10% best-paid employees. We also find that for the very top 0.1% of earners, who experienced particularly large earnings gains over the last decades, a larger share of earnings growth occurred within firms. However, because of their small number, the contribution of these very top earners to the overall increase in within-firm earnings inequality is small.

Keywords: Income inequality, pay inequality, between-firm inequality.

JEL Codes: E23, J21, J31

*Version: October 22, 2016. Special thanks to Gerald Ray at the Social Security Administration for his help and support. We thank our formal discussants Pat Kline, Ben Pugsley, Andre Shleifer, and seminar participants at the AEA, ASU, Berkeley, the White House CEA, Chicago, Dartmouth, Drexel, FRBs of Atlanta, New York, and Philadelphia, Harvard, Michigan, MIT, NBER, Northwestern, Princeton, Rand, Stanford, TNIT, UCLA, and Yale for helpful comments. Benjamin Smith and Brian Lucking provided superb research assistance. We are grateful to the National Science Foundation for generous funding. To combat alphabetical inequality, author names have been randomly ordered.

†Social Security Administration, jae.song@ssa.gov
‡Stanford University; djprice@stanford.edu
§University of Minnesota, FRB of Minneapolis, and NBER; guvenen@umn.edu
¶Stanford University, NBER, and SIEPR; nbloom@stanford.edu
kUCLA and NBER; tvwachter@econ.ucla.edu
1 Introduction

The dramatic rise in U.S. earnings inequality from the 1970s to today has been well documented (see Katz and Autor (1999) and Acemoglu and Autor (2011) for detailed reviews). It is well known that the change in inequality at the bottom (or below the median of the distribution) has been “episodic”—expanding in the 1980s and subsequently contracting and plateauing—whereas the rise in inequality above the median (all the way up to the very top earners) has been persistent throughout this period (Piketty and Saez (2003) and Autor et al. (2008)). An enormous body of theoretical and empirical research has been conducted over the past two decades in an attempt to understand the causes of these trends. Until recently, the analysis of the role of employers has been absent from this literature, chiefly because of the lack of a comprehensive, matched employer-employee data set in the United States that covers the period of rising inequality.

A long literature in economics has recognized that some firms pay workers with similar skills more than others (e.g., Slichter (1950), Dickens and Katz (1987), Krueger and Summers (1988), Van Reenen (1996)). Controlling for differences in the composition of observed and unobserved worker characteristics between firms, an increasing number of studies has shown that heterogeneity in firm pay contributes substantially to the distribution of earnings (e.g., Abowd et al. (1999), Goux and Maurin (1999), Abowd et al. (2002), Gruetter and Lalive (2009), Holzer et al. (2011)). A key question is to what extent the differences in firm pay have widened, and to what extent can this explain the observed increases in earnings inequality. In a recent paper, Card et al. (2013) have shown that a rise in the dispersion of firm pay has contributed substantially to recent increases in wage inequality in Germany. They also show that inequality rose in equal measure because of large changes in worker composition—high-wage workers became increasingly likely to work in high-wage firms (i.e., assortative matching increased), and high-wage workers became increasingly likely to work with each others (i.e., segregation rose).

Similar phenomena of changes in firm pay and changes in worker composition could also explain some of the shifts in inequality in the United States, which has experienced a stronger and more persistent increase in inequality than has Germany (as well as most other continental European countries). Indeed, as we discuss below, many of the mechanisms considered in the U.S. literature on inequality have potential implications for the contribution of firms and worker sorting to inequality, but these have not been evaluated so far. The firm dimension is also particularly interesting because it may help
us to better understand the rise in earnings at the very top, which many attribute to an increase in executive pay, a within-firm phenomenon. Recent findings in Barth et al. (2014) suggest that pay differences among firms and changes in worker sorting may indeed play an important role in understanding U.S. earnings inequality. Using data from several U.S. states from 1992 to 2007, they document an important rise in the variance of earnings between establishments, which they partly attribute to a change in the composition of observable worker characteristics.

In this paper, we study the contribution of firms and the role of worker flows in the rise of earnings inequality in the United States using a longitudinal data set covering both workers and firms for the entire U.S. labor market from 1978 to 2013. Our data set has several key advantages for studying firms and inequality: it is the only U.S. data set covering 100% of workers and firms for the entire period of the rise in inequality, it has uncapped W-2 earnings capturing a large share of earnings even at the very top, it has no lower earnings limit, and it has information on firms rather than establishments. Using this data set, in a first step we analyze the overall contribution of a rise in the variance of earnings between firms in explaining the evolution of U.S. earnings inequality from 1978 to today.

Our first main result is that the rise in the dispersion between firms in their average pay (the average of annual earnings for all workers in a given firm) accounts for the majority—two-thirds or more depending on the measure—of the increase in total earnings inequality. For example, examining one measure of inequality—the variance of log annual earnings—we show that the 19 log point increase in total variance between 1981 and 2013 is driven by a 13 point increase in the between-firm component and a 6 point increase within firms. This between-firm component captures all firm-related changes in inequality, including changes in firm pay as well as changes in sorting and segregation. Using our worker data, we show that this finding holds for changes in earnings in all but the very top percentile of the distribution. The importance of between-firm inequality in pay and worker characteristics explaining overall inequality trends is also seen in very fine industry, location, and demographic subsets of the economy, and is robust to different measures of inequality.¹

In a second step, we exploit the longitudinal nature of our data to assess to what extent the firm component we document is due to increasing differences in firm average pay or changing worker composition. We follow the modeling approach of Abowd et al.

¹For example, it holds true for the sample of continuing firms only and using five-year average measures of earnings.
(1999) [AKM] and Card et al. (2013) [CHK] to account for differences in both observed and typically unobserved permanent worker composition. Using this approach, we distinguish between two important drivers of inequality underlying our descriptive results: increases in the variance of firm pay, holding worker composition constant; and increases in assortative matching between high-wage workers to high-wage firms (which we will refer to as “sorting”). In addition, our data allow us to isolate the extent to which similar workers are increasingly likely to work together, which we will refer to as segregation. Although a rise in segregation by itself does not raise earnings inequality (because of a corresponding reduction in within-firm inequality), it leads to a higher contribution of firms in explaining earnings dispersion and could reflect important underlying economic forces.

Our second main finding is that the rise in between-firm inequality (in average pay) can be completely explained by changes in the composition of workers between firms. Taken together, increases in sorting (i.e., assortative matching of high-wage workers to high-wage firms) and segregation (i.e., the propensity of workers with similar mean wages to work in the same firm) explain the entire increase in between-firm inequality in our data. In contrast, the dispersion of actual firm pay net of worker characteristics was flat overall and, if anything, fell early during our sample period. An above-average increase in sorting and segregation can also entirely explain the substantially larger between-firm component in the variance of earnings we find for smaller firms. Although we find some interesting patterns of sorting between major industries, the majority of the rise in sorting and the rise in within-firm inequality occurs within industries.

Our third result is that the 31% of the increase in the total variance of annual earnings that occurs within firms comes mainly from large firms. The increase in the within-firm variance of log earnings in firms with 10,000+ employees (a group comprising 750 firms that employ about 23% of U.S. workers in 2013) is 58% between firms and 42% within firms (whereas the change in the variance of log earnings in firms with 20 to 1,000 workers is 92% between and 8% within firms). This rise in within-firm inequality in large firms comes from substantial changes at the bottom and the top of the within-firm earnings distribution. For example, between 1981 and 2013, median workers within 10,000+ employee firms saw their earnings fall by an average of 7%, those at the 10th percentile saw an average drop of 17%, whereas those at the 90th percentile saw an average rise of 11%. Overall, we calculate that the bottom half of the distribution is responsible for 35% of the rise in dispersion from 1981 to 2013. Changes in the 90th percentile and above explain about 45% of the secular rise in dispersion.
We also find that in these largest firms, the top 50 or so managers have seen substantial pay increases. For example, the average 50th highest-paid manager in large firms has seen real earnings rise by 47% between 1981 and 2013, while the average top-paid employee (presumably the chief executive officer) has seen real earnings rise by 137% over the same period. However, because there are few of these top-paid employees relative to the size of employment at these large firms (about 35,000 of them versus about 20 million total employees in these firms), we find that the impact of this rising top executive pay on explaining the increase in the variance in earnings is small. For example, the top 50 employees account for about 3% of the total increase in the within-firm dispersion of earnings from 1981 to 2013 at 10,000+ employee firms, whereas the top 5 employees account for less than 1% of the increase.

To summarize, our findings imply that the large rise in earnings inequality in the United States can be decomposed into three equally important forces—a rise in the dispersion of earnings within firms, a rise in the assortative matching of workers to firms, and a rise in the segregation of similar workers between firms. These findings highlight several potential mechanisms underlying rising earnings inequality. For example, it has long been hypothesized that persistent differences in firm pay reflect rent sharing (e.g., Dickens and Katz (1987), Katz and Summers (1989), Abowd et al. (1999)). Our finding of increasing assortative matching suggests that the distribution of rents may have become increasingly skewed, with an increasing share going to high-wage workers. Alternatively, these patterns could be related to complementarities in production between high-wage workers and high-wage firms. Reduction in search frictions, perhaps due to a decline in the cost of information or rising labor market intermediation (Autor (2009)), may have improved the assignment of workers to firms. A rise in sorting could also reflect an increase in such complementarities, perhaps reflecting uneven adoption of technological innovations. However, as we discuss in the paper, this had to occur without a rise in the variance of worker-firm match components in pay, which we find to be stable in our statistical analysis. Segregation could also reflect firms’ responses to a rising dispersion of skills (e.g., Kremer and Maskin (1996), Acemoglu (1999)), a rise in domestic outsourcing (Abraham and Taylor (1996), Dube and Kaplan (2010)), or a rise in the use of temporary workers (Segal and Sullivan (1997)).

Finally, our results are consistent with a substantial literature documenting that technological changes have increased inequality by shifting the demand for different skill groups (e.g., Autor et al. (2003), Acemoglu et al. (2007), Acemoglu and Autor (2011)). Both the substantial rise in within-firm inequality that we find, as well as some of the
changes in worker composition between firms, may reflect changes in market-level skill prices. In addition, the reduction in earnings for low-wage workers within large firms that we document corroborates the view that low-wage workers may have experienced a decline in access to high-paying jobs for institutional reasons, such as a decline in unionization or a change in company culture (e.g., Katz and Autor (1999)).

This paper is related to a large literature on inequality and, in particular, to a series of studies documenting that the variance of firm-specific pay explains an increasing share of total earnings inequality in a range of countries, including the United Kingdom (Faggio et al. (2007), Mueller et al. (2015)), Germany (Card et al. (2013)), Sweden (Häkanson et al. (2015)), and Brazil (Helpman et al. (2015), Alvarez et al. (2015)). In the United States, Davis and Haltiwanger (1991) were among the first to draw attention to the fact that rising inequality among workers was closely mirrored in rising inequality among establishments. However, these papers lacked data on wages within firms, which limited the scope of their analysis to between-firm data. The earlier finding was confirmed by Barth et al. (2014), who also find that a large share (about two-thirds in their analysis) of the rise in earnings inequality can be attributed to the rise in between-establishment inequality, concentrating on the period 1992 to 2007 for which they have both worker and establishment data for a subset of U.S. states. Our matched worker-firm data include information back to the 1970s and post-2007 to the Great Recession for all workers in the United States. As a result, we can examine the contribution of firms throughout the entire earnings distribution—including for the top end of the distribution that has attracted a lot of attention—consistently for the entire period of key changes in inequality.

A substantially smaller but growing literature has linked increases in between-firm inequality to changes in worker composition. Barth et al. (2014), Häkanson et al. (2015), Alvarez et al. (2015), and Card et al. (2013) document that changes in observable worker characteristics can account for an important share of the rise in the between-firm component in earnings inequality. Our approach follows that of Card et al. (2013), who use AKM’s method and find that changes in unobservable worker characteristics across firms can explain an important part of rising earnings inequality in Germany. We are the only paper that implements the AKM methodology for the entire U.S. labor market, which allows us to document the role of sorting and segregation for the full relevant period of increasing inequality.² Barth et al. (2014) and Card et al. (2013) also note the important distinction between sorting and segregation and document its importance. Direct evi-

²Preliminary work by Abowd et al. (2016) also uses the Longitudinal Employer-Household Dynamics data to analyze trends in inequality within and between firms, and finds similar results.
idence on the role of occupational segregation across industries and firms in the United States that is consistent with our findings is provided by Kremer and Maskin (1996) and Handwerker (2015), respectively.

Our results also speak to studies analyzing the sources of earnings inequality at the very top of the earnings distribution. Absent data on the distribution of wages within firms, a popular hypothesis has been that inequality at the very top of firms’ pay distribution is a driving force leading to an increase in overall inequality (e.g., Piketty (2013), Mishel and Sabadish (2014)), based on the earnings of about the top 5 earners within each firm from the Execucomp data. Our analysis of the entire wage distribution shows that these large pay increases within firms have been enjoyed by a broader set of managers—the top 50 or so managers—particularly in very large firms (10,000+ employees). In contrast, the top-paid employees—e.g., the top 1% of earners—at firms with less than 10,000 employees have seen their earnings rise more in line with the rise of the average earnings at their firm. Consequently, the contribution of top executives to the rise in overall inequality during this period was limited.

The paper is organized as follows. Section 2 describes the data set and the construction of the matched employer-employee data set, presents summary statistics from the sample, and discusses the methodology. Section 3 presents the main results. Section 4 decomposes the change in earnings inequality in components related to changes in firm average earnings, worker sorting, and worker segregation. Section 5 provides additional discussion on the sources of increases in within- and between-firm inequality, and Section 6 concludes.

2 Data

The main source of data used in this paper is the Master Earnings File (MEF), which is a confidential database compiled and maintained by the U.S. Social Security Administration (SSA). The MEF contains earnings records for every individual that has ever been issued a U.S. Social Security number. In addition to basic demographic information (sex, race, date of birth, etc.), the MEF contains labor earnings information for every year from 1978 to (as of this writing) 2013. Earnings data in the MEF are based on Box 1 of Form W-2, which is sent directly from employers to the SSA. Data from Box 1 are uncapped and include wages and salaries, bonuses, tips, exercised stock options, the dollar value of vested restricted stock units, and other sources of income deemed
as remuneration for labor services by the U.S. Internal Revenue Service. Because of potential measurement issues prior to 1981 (see Guvenen et al. (2014a)), we start most of our analysis from 1981, although results back to 1978 look similar. All earnings are converted to 2013 real values using the personal consumption expenditures (PCE) deflator.

Because earnings data are based on the W-2 form, the data set includes one record for each individual, for each firm they worked in, for each year. Crucially for our purposes, the MEF also contains a unique employer identification number (EIN) for each W-2 earnings record. Because the MEF covers the entire U.S. population and has EIN records for each job of each worker, we can use worker-side information to construct firm-level variables. In particular, we assign all workers who received wage earnings from the same EIN in a given year to that firm. Workers who hold multiple jobs in the same year are linked to the firm providing their largest source of earnings for the year. The resulting matched employer-employee data set contains information for each firm on total employment, wage bill, and earnings distribution, as well as the firm’s gender, age, and job tenure composition. Since we do not have information on hours or weeks worked, we measure individual annual earnings (or their total wage bill) rather than wage rates. As discussed in Subsection 2.3, we only include workers earning above a minimum threshold to minimize the effect of variation in hours worked.

In Figure 1a we plot the income distribution in real terms in 1981 and 2013. Looking at 2013, we observe a strikingly wide distribution of individual labor income—ranging from about $9,800 a year at the 10th percentile, to $36,000 at the median, $104,000 at the 90th percentile, and $316,000 at the 99th percentile. These figures are somewhat lower than data on earned income from, for example, Piketty and Saez (2003), primarily because they are based on individual (rather than household) values; see Figure A.18 in Appendix A. Comparing the 1981 and 2013 distributions, we can also see the increase in inequality as the 2013 distribution is increasingly pulling away from the 1981 distribution in the upper income percentiles, most notably for the top 1% in Figure 1b. These patterns have been studied extensively in the literature on earnings inequality. Here, we focus on the role of employers in accounting for these changes.

---

3The MEF has previously been used by, among others, Davis and Von Wachter (2011) and Guvenen et al. (2014b), who describe further details of the data set. Kopczuk et al. (2010) use the 1% Continuous Work History Subsample (CWHS) extract of SSA data to conduct an extensive analysis of long-run trends in mobility.
2.1 What Is a Firm?

Throughout the paper, we use employer identification numbers (EINs) as the boundary of a firm. The EIN is the level at which companies file their tax returns with the IRS, so it reflects a distinct corporate unit for tax (and therefore accounting) purposes. Government agencies, such as the Bureau of Labor Statistics, commonly use EINs to define firms.\(^4\) They are also often used in research on firms based on administrative data. An EIN is not always the same, however, as the ultimate parent firm. Typically, this is because large firms file taxes at a slightly lower level than the ultimate parent firm.\(^5\) Although it is unclear what level of aggregation is appropriate in order to define a “firm,” we follow much of the existing literature and view the EIN as a sensible concept reflecting a unit of tax and financial accounting. An EIN is a concept distinct from an


\(^5\)For example, the 4,233 New York Stock Exchange publicly listed firms in the Dunn & Bradstreet database report operating 13,377 EINs, or an average of 3.2 EINs each. For example, according to Dunn & Bradstreet, Walmart operates an EIN called “Walmart Stores,” which operates the domestic retail stores, with different EINs for the Supercenter, Neighborhood Market, Sam’s Club, and On-line divisions. As another example, Stanford University has four EINs: the university, the bookstore, the main hospital, and the children’s hospitals.
“establishment,” which typically represents a single geographic production location and is another commonly used unit of analysis to study the behavior of “firms” (e.g., this is the definition used by Barth et al. (2014), who study inequality using U.S. Census data). Around 30 million U.S. establishments in the Longitudinal Business Database in 2012 are owned by around 6 million EIN firms, so an establishment is a more disaggregated concept. As Abowd et al. (2016) show, 84% of the increase in cross-establishment inequality can be accounted for by firms, so firms are an appropriate unit of analysis.

2.2 Benchmarking the Master Earnings File against Other Data Sets

We found our data matches key moments from standard nationally representative data sets quite well. In particular, when compared to the Current Population Survey (CPS) the SSA data matches the changes in the variance of log annual earnings quite closely; see Figure A.2.\textsuperscript{6} We also checked a range of other statistics. For example, aggregating wages and salaries from all W-2 records over all individuals in the MEF yields a total wage bill of $6.8 trillion in 2013. The corresponding figure from the national income and product accounts (NIPAs) is $7.1 trillion, so these numbers are very close; see Figure A.1a for the two series over time. While the level of employment is higher in the MEF than in the CPS, trend in the total number of individuals in the MEF who received W-2 income in a given year (our measure of total employment) also closely tracks total employment in the CPS (see Figure A.1b).\textsuperscript{7} There are 6.1 million unique firms (EINs) in the MEF in 2013, each associated with at least one employee. This number is slightly higher than the 5.8 million firms (with employees) identified by the Census Bureau’s Statistics of U.S. Businesses data set. In addition, as shown in Appendix Figure A.1c, the trends in each of these data sets are similar over time (at least since 1988, when the Census data begins).

\textsuperscript{6}Although the change in variance is comparable, the level of variance is higher in SSA data. This may be because SSA data is not top-coded, and because those with lower incomes may not report them in the CPS. For reference, Figure A.3 shows the cumulative distribution of earnings in the CPS data, which is comparable to Figure 1a for SSA data.

\textsuperscript{7}In 2013, for example, the MEF measure contains 155 million workers, while the CPS indicated that, on average, 144 million individuals were employed at any given time. The difference is likely because the CPS is a point-in-time estimate; if people cycle in and out of employment, they may be missed in the CPS data but will be included in the MEF (which is an aggregate measure over the year). Furthermore, the CPS excludes the institutionalized population, whereas the MEF includes them.
2.3 Baseline Sample

For our descriptive analysis in Section 3, we restrict our baseline sample to individuals aged 20 to 60 who work full-time, where “full-time” is defined as earning at least that year’s minimum wage for one quarter full-time (so for 2013, 13 weeks for 40 hours at $7.25 per hour, or $3,770). These restrictions reduce the effect on our results of individuals who are not strongly attached to the labor market. We also restrict to firms (and workers in firms) with 20+ employees to help ensure that within-firm statistics are meaningful. We exclude firms (and workers in firms) in the government or educational sectors, because organizations in those sectors are schools and government agencies rather than what economists think of as firms. This yields a sample of, on average, 72.6 million workers and 477,000 firms per year, rising from 55.5 million and 371,000 in 1981 to 85.2 million and 517,000 in 2013, respectively. As we show in Appendix B, none of our results are sensitive to these assumptions—the results look similar using all ages, all firm sizes, all industries (Figure A.13), and minimum earnings thresholds up to full-time (2,080 hours) at minimum wage (Figure A.14). Some statistics describing the sample are shown in Table 1. More details about the data procedures are discussed in Appendix B.

3 Inequality within and between Firms

3.1 Rising between-Firm Inequality

The first key result in this paper—that rising earnings inequality in the United States reflects rising dispersion in average earnings between firms—is most easily seen graphically in a number of ways. We first examine a decomposition of variances over time. We then look at earnings percentiles, which is an approach similar to examining yearly changes in inequality over time, but focuses on particular key percentiles of the earnings distribution. Finally, we examine the long difference in earnings between 1981 and 2013, but we do this for each percentile by worker and their firms, providing rich cross-sectional analysis but across one time period. As will become clear, all three approaches show a similar result: rising inequality is to an important degree a between-firm phenomenon. Until Section 4, we make no distinction between changes in firm pay and changes in worker composition, both of which could be driving our findings in the descriptive analysis.
### Table 1 – Percentiles of various statistics from the data

<table>
<thead>
<tr>
<th>Year</th>
<th>Group</th>
<th>Statistic</th>
<th>10%ile</th>
<th>25%ile</th>
<th>50%ile</th>
<th>75%ile</th>
<th>90%tile</th>
</tr>
</thead>
<tbody>
<tr>
<td>1981</td>
<td>Firm</td>
<td>Earnings (Unwgt)</td>
<td>12.6</td>
<td>16.6</td>
<td>23.8</td>
<td>32.5</td>
<td>41.9</td>
</tr>
<tr>
<td>1981</td>
<td>Firm</td>
<td>Earnings (Wgted)</td>
<td>15.2</td>
<td>21.5</td>
<td>30.6</td>
<td>43.2</td>
<td>52.1</td>
</tr>
<tr>
<td>1981</td>
<td>Firm</td>
<td>Employees</td>
<td>22</td>
<td>26</td>
<td>38</td>
<td>73</td>
<td>169</td>
</tr>
<tr>
<td>1981</td>
<td>Indiv.</td>
<td>Earnings</td>
<td>9.46</td>
<td>18.2</td>
<td>31.9</td>
<td>51.7</td>
<td>73.8</td>
</tr>
<tr>
<td>1981</td>
<td>Indiv.</td>
<td>Earnings/Firm Avg</td>
<td>0.43</td>
<td>0.724</td>
<td>1.05</td>
<td>1.45</td>
<td>2.06</td>
</tr>
<tr>
<td>1981</td>
<td>Indiv.</td>
<td>Employees</td>
<td>42</td>
<td>127</td>
<td>1153</td>
<td>12418</td>
<td>62718</td>
</tr>
<tr>
<td>2013</td>
<td>Firm</td>
<td>Earnings (Unwgt)</td>
<td>13.8</td>
<td>19.3</td>
<td>30.5</td>
<td>43.8</td>
<td>61.4</td>
</tr>
<tr>
<td>2013</td>
<td>Firm</td>
<td>Earnings (Wgted)</td>
<td>15.3</td>
<td>21.4</td>
<td>35.8</td>
<td>52.1</td>
<td>73.6</td>
</tr>
<tr>
<td>2013</td>
<td>Firm</td>
<td>Employees</td>
<td>22</td>
<td>26</td>
<td>39</td>
<td>79</td>
<td>189</td>
</tr>
<tr>
<td>2013</td>
<td>Indiv.</td>
<td>Earnings</td>
<td>9.82</td>
<td>19.2</td>
<td>36</td>
<td>63.2</td>
<td>104</td>
</tr>
<tr>
<td>2013</td>
<td>Indiv.</td>
<td>Earnings/Firm Avg</td>
<td>0.421</td>
<td>0.681</td>
<td>1.03</td>
<td>1.5</td>
<td>2.22</td>
</tr>
<tr>
<td>2013</td>
<td>Indiv.</td>
<td>Employees</td>
<td>45</td>
<td>157</td>
<td>1381</td>
<td>14197</td>
<td>78757</td>
</tr>
</tbody>
</table>

Notes: Values indicate various percentiles for the data for individuals or firms. All dollar values are in thousands and are adjusted for inflation using the PCE deflator. Only firms and individuals in firms with at least 20 employees are included. Firm statistics are based on mean earnings at firms and are either unweighted or weighted by number of employees, as indicated. Only full-time individuals aged 20 to 60 are included in all statistics, where full-time is defined as earning the equivalent of minimum wage for 40 hours per week in 13 weeks. Individuals and firms in public administration or educational services are not included.

#### 3.1.1 Simple Variance Decomposition

One straightforward analysis is to decompose the variance of earnings into a within- and between-firm component. In particular, let $y_{i,j}^t$ be the log earnings of worker $i$ employed by firm $j$ in period $t$.\(^8\) This can be broken down into two components:

$$y_{i,j}^t = \bar{y}_j^t + [y_{i,j}^t - \bar{y}_j^t], \quad (1)$$

where $\bar{y}_j^t$ is the average wage earnings paid by firm $j$, enabling us to simply define the decomposition of variance:

$$\text{var}_i(y_{i,j}^t) = \text{var}_j(\bar{y}_j^t) + \text{var}_i(y_{i,j}^t | i \in j). \quad (2)$$

This equation provides a straightforward way to decompose the total earnings dispersion in the economy into (i) the between-firm dispersion in average earnings paid by firms and (ii) the within-firm dispersion.

---

\(^8\)For notational convenience, we suppress the dependence of the subscript $j$ on worker $i$. 

12
Figure 2 – Decomposition of Variance in Annual Earnings within and between Firms: All, Smaller, and Larger Firms

(A) Overall decomposition

(B) Workers at Firms with 20 to 10,000 employees

(C) Workers at Firms with 10,000+ employees

Notes: See variance decomposition in equation (2). Only firms and individuals in firms with at least 20 employees are included. Only full-time individuals aged 20 to 60 are included in all statistics, where full-time is defined as earning the equivalent of minimum wage for 40 hours per week in 13 weeks. Individuals and firms in public administration or educational services are not included. Firm variance is calculated using mean log earnings and weighted by number of employees. Within-firm variance is calculated based on the difference between individual log earnings and firm mean log earnings.

each firm and (ii) the within-firm dispersion in pay weighted by the employment share of each firm. The components of equation (2) are plotted separately in Figure 2a. We find that the overall variance of log earnings has risen by 19 log points between 1981 and 2013, very similar to the results in Autor et al. (2008). Examining the within- and between-firm components of this increase, we see that 13 log points of this variance
arise between firms and 6 log points within firms, so that 69% of the overall increase in inequality evaluated on this metric is a between-firm phenomenon.

### 3.1.2 Tracking Select Percentiles

Another way to examine income inequality over time is by tracking the evolution of select income percentiles. In Figure 3a we plot the change in average log earnings within the 99th, 90th, 50th, and 25th percentiles, revealing the well-known result that earnings inequality has increased since 1981, with higher percentiles enjoying substantially larger earnings growth. Since our sample covers around 70 million workers, each one of these percentiles contains around 0.7 million workers per year.

In Figure 3b we plot the change in average log earnings in the firms for each of these individual earnings percentiles. So, for example, the 99th percentile point for Figure 3b reports the increase in average earnings for the colleagues of the individuals in the 99th percentile line of Figure 3a.\(^9\) Finally, in Figure 3c we report the relative change in the earnings of individuals compared with their colleagues and reveal a set of flat percentiles. In short, while individuals have seen a large increase in pay inequality across their earnings percentiles since 1981, this increase has been tracked very closely by the earnings of their colleagues. So, for example, although the 99th percentile has seen real earnings increase by 51 log points between 1981 and 2013, the log earnings of their colleagues in the 99th percentile have increased by an average of 49 log points; thus, these individuals saw only a 2 log point increase in earnings relative to their colleagues.

We should also note that this measure does not use any of the panel structure of the data; individuals in the 50th percentile in 1981 are almost certainly different from those in the 50th percentile in 2013. In Section 4, we will undertake a type of panel analysis pioneered by Abowd et al. (1999) and reveal that not only has inequality increased in the cross section, but the inequality of the persistent worker component of earnings has also shown a substantial increase.

### 3.1.3 Inequality across Percentiles

Because inequality is a concept about the entire income distribution, the simple summary statistics (such as the variance and select percentiles) reported above can mask interesting and important variation hidden in various parts of the distribution. A more detailed look is provided by a graphical construct popularized by Juhn et al. (1993) and

---

\(^9\)That is, the line shows \(\delta_{q}^{firm} = E[\bar{y}_{2013} | i \in Q_{2013,q}] - E[\bar{y}_{1981} | i \in Q_{1981,q}]\), where \(Q_{t,q}\) is the set of individuals in the \(q\)th percentile in year \(t\), and \(j\) refers to the employer of worker \(i\).
Figure 3 – Change in percentiles of annual earnings within and between firms relative to 1981

(A) Individuals

Change Since 1981


Year

Notes: Only firms and individuals in firms with at least 20 employees are included. Only full-time individuals aged 20 to 60 are included in all statistics, where full-time is defined as earning the equivalent of minimum wage for 40 hours per week in 13 weeks. Individuals and firms in public administration or educational services are not included. Firm statistics are based on the average of mean log earnings at the firms for individuals in that percentile of earnings in each year. Data on individuals/their firms are based on individual log earnings minus firm mean log earnings for individuals in that percentile of earnings in each year. All values are adjusted for inflation using the PCE price index.

has been used extensively since then. We now use a variation of this graphical construct in three steps.

First, we start with individuals in the baseline sample in 1981 and group them into percentile bins on the basis of their income in 1981. Then we calculate the average of log real earnings (in 2013 dollars) for each percentile bin. Let $P_t x$ denote this average
for percentile bin $x$ in year $t$. We then repeat the same procedure for 2013 (now for individuals who satisfy the sample selection in that year). The blue line marked with diamonds (labeled “Indv Total Earnings”) in Figure 4 plots $P_{2013,x} - P_{1981,x}$ for all percentile groups $x = 1, 2, ..., 99, 100$ against the percentile number $y$ on the horizontal axis. So, for example, we see that between 1980 and 2013, the 50th percentile of earnings has increased by 12 log points (13%) from about $31,500 to $35,600. The upward slope of the individual line highlights the rise in individual earnings inequality—earnings at higher percentiles have risen at a faster rate, and this rise grows steadily as you move up the income percentiles.

For the red line marked with circles (labeled “Avg of Log Earnings at Firm”), we put individuals into percentile bins based on their own wage earnings in 1981—just as we did for the “Individuals” line above—but for each percentile bin, we calculate the average of the mean log real earnings at each individual’s employer (or firm). We repeat the same procedure for 2013. For example, in 1981, individuals in the 50th percentile of individual earnings were employed in firms with average log mean real earnings of 10.30 (corresponding to about $29,900); in 2013, individuals in the 50th percentile were employed in firms with average log mean real earnings of 10.45 (corresponding to about $34,700). The difference of 0.15 (i.e., 15 log points) is plotted on the graph at the 50th
percentile.

Finally, the green line marked with squares (labeled “Indv Earnings/Firm Average”) is based on the residual earnings measure, $y^{i,j}_t - \bar{y}^j_t$. Specifically, we compute the average of $y^{i,j}_t - \bar{y}^j_t$ across all workers within a percentile in each year.\(^{10}\) We then plot the change in this statistic between 1981 and 2013. For example, in 1981, individuals in the 50th percentile of individual earnings had average log earnings that were 0.05 higher than their firms’ mean earnings (corresponding to about 106% of their firms’ mean earnings). In 2013, individuals in the 50th percentile had average log earnings that were 0.03 higher than their firms’ mean earnings (corresponding to about 103% of their firms’ mean earnings). We plot the difference of −0.02 at the 50th percentile. Note that this “Individual/Firm” line will be mechanically equal to the difference between the “Individual” line and the “Firm” line.

For all these graphs, results should be interpreted similarly. A flat line indicates that inequality for that statistic has not changed over the time period, because the statistics for those at the top and the bottom have changed by the same amount. An upward-sloping line indicates that inequality has increased, because the statistic for those at the top has increased more than the statistic for those at the bottom; and by the same logic, a downward-sloping line indicates that inequality has decreased. This graphical construct thus allows us to detect changes in inequality that might be confined to one part of the earnings distribution and may not be very visible in broad inequality statistics.

Particular care should be given to the interpretation of the green (square symbol) “Individual/Firm” line. The level of this line indicates the extent to which a particular demographic group gains or loses relative to the firm average for all employees. When we examine the whole population or subsets of the population that, for each firm, include either everyone or no one at that firm, the green line’s weighted average taken over all percentiles must be zero. The key finding from this figure is that the green line is almost flat across all percentiles, indicating that within-firm inequality has remained nearly constant over the entire population of workers.

### 3.2 Robustness of Results

The results above show that, perhaps rather surprisingly, the majority of the increase in earnings inequality among workers is associated with a rise in the dispersion of mean earnings among firms and their employees. To investigate the robustness of these results,

\(^{10}\)Notice that in all likelihood, the workers we average over are employed in different firms, but each residual is computed with respect to a worker’s own employer.
we reran the analysis from Figure 4 within many subgroups and using many different definitions. The basic result—that between-firm inequality accounts for most of the inequality—remains true for each such analysis.

First, given the different trends in rents and amenities identified by Moretti (2013) and Diamond (2016), could this increase in between-firm inequality simply reflect regional variation? To investigate this question, we reran our analysis within each county and took the average (Figure A.2b) and within each census region (Figure A.5), finding very similar results. In case these trends reflect changes in demographics, Figure A.8 reports the graphs by age breakdowns (20s, 30s, 40s, and 50s) and by gender, with again the results looking broadly similar for each subgroup. Another possible driver could be variations by industry—perhaps differential trends arising from trade, technology, or other industry factors are driving the firm results (e.g., Autor et al. (2013) and Pierce and Schott (2016))? However, the results are also similar within broad industry SIC one-digit categories as shown in Figure A.9, and also on average within narrow SIC four-digit categories as shown in Figure A.10.

Another possible concern would be if the increase in earnings inequality within firms is driven by differences across establishments. Using the Census Longitudinal Business Database (LBD), which covers all establishments in the United States, we decompose the variance in average earnings differences across establishments into a between-firm and a within-firm component. We see in Figure A.6 that for the same sample as our main SSA analysis (firms with 20+ employees in all sectors excluding public and education), the increase in the variance of the average of log earnings across establishments has been 12 log points, with the bulk of this rise (10 log points) across firms. So in our overall sample, inequality is primarily a between-firm (rather than within-firm but between-establishment) phenomenon. Of course, this sample contains many smaller firms so it might be expected that the majority of the increase in inequality is across rather than within firms. But in Figure A.7, we examine the sample of establishments in firms with 10,000+ employees and find similarly that of the 15 log point increase in earnings inequality the large majority (11 log points) was between firms. Hence, examining the rising inequality across firms is capturing the large majority of the rising inequality across.

11 For age and gender graphs, to improve comparability between them and with the firm statistics, sorting into percentile bins is based on the overall population. Note that the interpretation of the “Individual Earnings/Firm Average” is different when we look at demographic subsets of the population, where we examine only a subset of workers from each firm. For these analyses, there is no presumption that the level for any group must have an average of zero; instead, we interpret the average level as the extent to which the group has gained or lost relative to the firm average weighted by group-specific employment.
workplaces in the United States.\textsuperscript{12}

We also experimented with different measures of the firm average earnings, using the log of average earnings (rather than the average of log earnings), firm median earnings, and the average log earnings among only those in the bottom 95%, and calculating firm average log earnings leaving out the individual themselves; once again we find very similar results (see Figure A.11). We examined a panel of continuing firms in case our results were being driven by the selection of firms, and once again, we see very similar results (see Figure A.12). We varied our definition of full-time earnings from 520 hours at the 2013 minimum wage to 260, 1040, and 2080 hours and again find broadly similar results (see Figure A.14). Finally, in Figure A.15 we compare changes in five-year (rather than one-year) earnings in case the results were being driven by changes in temporary (rather than permanent) earnings, but we see very similar results.

We also considered other robustness issues around health care and self-employment income. On health care, perhaps rising firm earnings inequality is offset by an increase in the generosity of firm health care insurance that, as a flat entitlement to all employees, provides a progressive compensation component. In fact, as Burkhauser and Simon (2010) show, \textit{employer-provided} (but not government) health insurance is about as unequally distributed as earnings among the bottom eight income deciles. Kaestner and Lubotsky (2016) show that \textit{employer-provided} health insurance actually increases inequality. Higher-paid employees are more likely to be in firms offering generous health care packages, have higher firm coverage rates, pay lower premiums, and are more likely to enroll.\textsuperscript{13}

Regarding self-employment, the IRS Statistics of Income reports that in 2012, 16.5% of individuals reported self-employment income on Schedule C and 1099 forms, while it accounted for only 3.2% of all income, most of which is concentrated in employees of smaller firms. Hence, in our 20+ employee sample, self-employment income is too small to play a major role in shaping inequality.

So overall, the basic result that the majority of increasing inequality is a related to changes in firm pay or changes in worker composition seems to be broadly robust. The two groups that are partial exceptions to this are the top 1% of earner and very large employers, to which we turn next.

\textsuperscript{12}Barth et al. (2014) and Abowd et al. (2016) come to a similar conclusion.

\textsuperscript{13}The part of health care that has reduced inequality is Medicaid and Medicare, programs that are strongly progressive and have increased in generosity (Burkhauser and Simon, 2010). However, since this part of health care is independent of the employee-firm match, this does not influence our analysis.
3.3 The Top 1%

Much of the recent policy and media attention around inequality has focused on the top 1% of earners, following in particular the early work of Piketty and Saez (2003). We find that the role of firms in increasing earnings inequality in the top 1% (and particularly the top 0.5% and the top 0.1%) is different than the rest of the earnings distribution. To show this, Figure 5 plots the cross-sectional percentiles graph for just the top 1%, breaking this into 100 subdivisions of 0.01% each. (Our baseline sample covers about 55 million workers in 1981 and 85 million workers in 2013, for an average of 70 million over the sample period, so each 0.01% represents about 7,000 people on average.)

We see in Figure 5 that up until about the 99.5% point—which is an earnings threshold of around $450,000 in 2013 (see Figure 1b)—increases in individual earnings from 1981 to 2013 within each percentile point have been matched almost fully by the increases in earnings of their firms. However, in the top 0.5% and particularly the top 0.1%, there is such a steep increase in earnings between 1981 and 2013 that these rises have outpaced those of their colleagues. For example, the 99.95th percentile reveals individual earnings growth of 102 log points (178%), while the firms these employees work for have increased...
their average pay by 73 log points (107%), generating a 30 log point gap.\footnote{Most of this divergence between top workers and their firms occurred between 1981 and about 1988; since then, earnings of even those at the top of the top 1% have risen similarly to their firms’ earnings.}

Thus, a group of about 70,000 people representing about the top 0.1% of earners has seen substantial pay increases over and above those of their colleagues. This group likely includes the chief executive officers of some very large companies, but also a far wider group of individuals including physicians, finance professionals, lawyers, engineers, among others (Guvenen et al. (2014a)).

### 3.4 Inequality Among Employees by Firm Size

Another exception to our main findings is that the rise in earnings inequality among workers employed at very large firms is driven to a large degree by increases in within-firm inequality. Figures 2b and 2c plot the decomposition of variance for firms with less than and greater than 10,000 workers. We see in Figure 2b that in firms with less than 10k workers—which contain over 70% of employees and over 99% of firms—inequality is almost entirely (84%) due to between-firm variation. In comparison, increases in inequality in the 10k+ worker firms—which account for about 30% of employees and only about 700 firms—is still mostly (58%) between firms but also has a large (42%) within-firm component. We also find that the rise in within-firm earnings inequality among the top 1% of earners discussed in Section 3.3 is more pronounced at very large employers. We discuss the phenomenon of rising within-firm inequality among very large employers and its potential sources in more depth in Sections 4 and especially Section 5.

### 4 The Role of Worker Sorting and Segregation

The rise in the dispersion in average earnings between employers we document in Section 3 could come from two different sources. First, a “widening firm premium” story: firms may be increasingly unequal in their earnings because some firms had become economic “winners” and are sharing the increased profits with their workers, whereas other “loser” firms are not. Second, a “worker composition” story: high-wage workers may be increasingly sorted into high-wage firms, or workers may be increasingly segregating among firms (so that high-ability workers are clustering in some firms and low-ability workers in others). As we show below, the phenomena of worker sorting and worker segregation appear to jointly account for almost the entire increase in between-firm inequality we documented in Section 3.
4.1 Econometric Model of Worker and Firm Effects

To analyze the worker and firm movements in earnings we follow the Card et al. (2013) [henceforth CHK] implementation of the model introduced by Abowd et al. (1999) [henceforth AKM] and solved by Abowd et al. (2002). We will divide our time period into five seven-year periods, as discussed further below, and estimate a separate model for each period \( p \). The regression model we estimate in each period is

\[
y_{i,j}^{p} = \theta^{i,j} + X_{i}^{p} \beta^{p} + \psi^{j,p} + \epsilon_{i,j}^{p},
\]

where \( \theta^{i,j} \) captures earnings related to fixed worker characteristics (such as returns to formal schooling or to innate ability), \( \beta^{p} \) captures the effect of time-varying worker characteristics (in our case, a polynomial in age and year effects), and \( \psi^{j,p} \) captures persistent earnings differences related to firm \( j \) (such as sharing of rents or compensating differentials). The residual, \( \epsilon_{i,j}^{p} \), captures purely transitory earnings fluctuations. In addition, the residual will also contain any worker-firm specific (match) components in earnings, which we will denote by \( m^{i,j} \).

The AKM model has proven to be an empirically successful extension of the standard human capital earnings function and has developed into the workhorse model for incorporating firm components into traditional earnings regressions. We confirm that the model appears to summarize a range of key patterns in our data surprisingly well. Hence, despite well-known limitations we discuss below and in Appendix C, we believe that there is sufficient support for the model to treat it as a useful diagnostic device to better understand the patterns underlying the stark changes in the between-firm component over time.

The estimates of the parameters of the econometric model in equation (3) can be used to further decompose the within- and between-firm components of the variance. Ignoring time-varying worker characteristics \( X_{i}^{p} \beta^{p} \) for now and variation across periods (dropping superscript \( p \)), the firm-worker decomposition is

\[
\text{var}(y_{i,j}^{p}) = \text{cov}(\bar{\psi}^{j}, \psi^{j}) + \text{var}(\bar{\theta}^{j}) + \text{var}(\theta^{i} - \bar{\theta}^{j}) + \text{var}(\epsilon_{i,j}^{p}).
\]

To simplify notation, we leave the dependence of the identity of the firm on the worker implicit, such that \( j \equiv j(i) \). Note that while most of the literature uses the model to analyze daily or hourly wages, we follow an increasing number of papers that analyze earnings. We discuss the potential role of labor supply differences below.
where the moments in the between-firm component are weighted by the number of worker-years.\textsuperscript{16} Equation (4) shows how the between-firm component of the variance can be decomposed into three pieces: a part deriving from the variance of firm effects $\text{var}(\psi^j)$, a part deriving from the covariance of worker and firm effects, $\text{cov}(\overline{\theta}^j, \psi^j)$, and a part deriving from the variance of the average worker effect in each firm, $\text{var}(\overline{\theta}^j)$.

The first component is the “widening firm premium” part—perhaps because the variance of firm pay has increased. The second component reflects the “worker-sorting” story—high-paid workers are increasingly sorting into high-paying firms. The third part is the “worker segregation” story—lower- and higher-paid workers are segregating into different firms. Splitting the worker component into the sorting covariance and segregation variance terms allows us to better characterize the role of firms in accounting for earnings inequality, since sorting increases aggregate inequality whereas segregation does not.

4.2 Implementation of Regression Model Using SSA Data

We estimate equation (3) separately for five adjacent seven-year intervals beginning in 1980 and ending in 2013.\textsuperscript{17} As is well known, firm fixed effects are identified by workers moving between firms and hence can only be estimated relative to an omitted firm. Estimation of equation (3) is done on the largest set of firms connected by worker flows. We impose the same restrictions on the data as in our descriptive analysis, with three exceptions. To maximize the number of observations in the connected set, we do not impose a restriction on firm size and do not exclude the education and public sector. Because of limitations in computing power, we present worker and firm effects only for men in this version of the paper. All other restrictions, including imposing a minimum earnings threshold, are the same as described in Section 2.

Although our implementation of AKM follows CHK, an important difference is that we have data on annual earnings for all workers, not daily wages for full-time workers. This means that our estimates of worker and firm effects may capture systematic differences in labor supply between workers and firms.\textsuperscript{18} Given the nature of our data,

\textsuperscript{16}Note that one can rewrite the within-firm component as $\text{var}(\theta^i - \overline{\theta}^j) = E_j\{\text{var}(\theta^i|i \in j)\}$, that is, as the worker-weighted mean of the firm-specific variances of the worker effect (and similarly for $\text{var}(\epsilon^i_{i,j})$).

\textsuperscript{17}The choice of intervals trades off limitations in computational power and the desire to analyze changes in the variance over time with the sampling error in estimates of the worker and firm effects and the resulting bias in the variance and covariance terms, which depend on the number of movers between firms. We experimented with intervals up to ten years and found that our results did not change substantially.

\textsuperscript{18}In that sense, our implementation is comparable to Abowd et al. (2016), who implement this model
such differences can arise because of variation at both the intensive margin (i.e., hours worked) and the extensive margin (i.e., days worked in a year). In principle, these differences could affect the level and change of the moments in our variance decomposition. However, it is worth noting that under the plausible assumption that job moves occur randomly within a year, there is no mechanical reason why labor supply effects should introduce a bias into our estimates of firm effects.

We tried various ways to address the potential effects of systematic labor supply differences in our findings. We have experimented by imposing increasingly stringent lower earnings restrictions. Using retrospective data from the CPS, one can show that this approach tends to eliminate part-time or part-year workers. Our results are robust to variation in this restriction; see, for example, Figure A.14. Since our analysis based on the CPS also shows that more stringent earnings cutoffs eliminate low-wage full-time or full-year workers, we use a less stringent restriction in our main sample. CPS data do not reveal any trend in the aggregate variance of weekly hours worked or weeks per year worked over time. Given the robustness of our findings and the stability in trends in the variance of time worked, we are confident that our main results are mainly driven by changes in the variance of wages, not hours or days worked.

Estimating the model requires a set of identification assumptions, which given the prior literature on this, we do not discuss in detail in the paper and relegate to Appendix C. Since the estimated firm effects will capture any systematic differences in earnings of movers before and after the job move, to associate estimated firm effects with true underlying firm-specific differences in pay, we have to assume that conditional on worker and firm effects, job moves do not depend systematically on other components, in particular worker-firm specific job match effects (the conditional random mobility (CRM) assumption). After reviewing the evidence, we join an increasing number of papers whose results indicate that the AKM model can be estimated without too much systematic bias (e.g., AKM, CHK, and Abowd et al. (2016)).

In particular, we do not find a role of an increasing dispersion in worker-firm match using quarterly earnings.

19 For example, systematic differences in the propensity to take part-time jobs or to be unemployed would load onto the worker fixed effect. If firms offer different hours packages or offer seasonal work, this could load onto the firm effect as well. If high-hour workers (or stable workers) are increasingly sorted into high-hour firms (or stable firms), labor supply can also affect the nature of sorting. If job moves are partly triggered by changes in hours worked, labor supply effects could also contribute to a failure of the conditional random mobility (CRM) assumption.

20 If one compares the number of observations in our final sample with the number of workers, one obtains that the average worker is in the sample for about five of seven years in each period. This number is very similar to numbers reported by CHK (Table I) for full-time male workers in Germany.
effects in explaining rising inequality. To check whether adding a match-specific component would substantially increase the fit of the model, we indirectly included a match effect \( (m_{ij}) \) in the model. Although, not surprisingly, allowing for a match effect reduces the RMSE and increases the adjusted \( R^2 \), the standard deviation of match effects declines somewhat over time. Similarly, we also find that the goodness of fit of the model without a match component has increased over time from an \( R^2 \) of 74% (1980-1986) to an \( R^2 \) of 81% (2007-2013), driven by both a reduction in the root mean squared error (RMSE) and an increase in the variance of earnings. If the rise in the sorting of workers to firms that we find had resulted from an increasing role of complementarities (i.e., match effects), we would have expected the RMSE to rise and the goodness of fit of the model without match effects to decline over time (see Appendix Table C2).

Finally, if the model is correctly specified, on average workers changing from one firm to another should experience earnings changes corresponding to the estimated firm effects. To implement this comparison, we used our data to perform event-study analyses of the effect of job mobility on earnings akin to those shown in CHK (see their Figure VII). In Figure 6, we divided firms into quartiles according to their estimated firm effects and recorded the mean earnings of workers moving between the four firm-type classes in the years before and after the job change. On average, the patterns of earnings changes are approximately symmetric for switches between firm groups, and there are no signs of systematic earnings declines or increases before or after job changes, both of which are consistent with the CRM assumption. Overall, despite being an obvious abstraction from reality, we conclude that our model constitutes a useful tool for a better understanding of trends in earnings inequality in our data.

### 4.3 Decomposing the Change in the Variance of Earnings

Table 2 presents results for the variance decomposition of earnings (equation 4) for our five periods, as well as for the change from period 1 (1980-1986) to period 5 (2007-2013).

---

21 Since violations of the separability assumptions in the AKM model would likely cause large mean residuals for certain matches (say, where highly skilled workers are matched to low-wage establishments), we directly examined the distribution of average residuals by 100 cells of estimated firm and worker effects. For most cells, the mean residual is very small, below 0.02, and shows few systematic patterns (see Appendix Figure A.19).

22 To deal with the fact that we do not know the specific time of the move, we followed workers from two years before the year \( t \) in which we observe the move (i.e., from year \( t-2 \) to \( t-1 \)), to two years after the year succeeding the move (from year \( t+2 \) to \( t+3 \)). To try to further approximate the transition between “full-time” jobs, we only look at workers who remained at the firm in the two years before and two years after the move. Since we are following workers for six years, we adjust earnings for flexible time trends. In Appendix C, Figure A.20, we show a version of the figure in which the four firm classes are generated based on average earnings within the firm.

---

25
There are two key findings. First, in all periods, about half of the level of the variance of log annual earnings is explained by the variance of worker effects, which at 48% – 53% is by far the biggest component. The two other within-firm components, the variance of time-varying worker characteristics (7%–10%) and the variance of residuals (12% – 16%) together explain roughly another 20% of the variance. Firm-related components together (the variance of fixed effects, their correlation with worker effects and worker characteristics, and the variance of average worker characteristics across firms) explain about 20% to 25% of the level variance.

Second, although worker effects dominate in accounting for the level of the variance,
when examining the change in the variance over time, the firm-based components dominate. As shown in the final column of Table 2, consistent with our results in Section 3 the firm component explains 69.1% of the rise in the overall variance.²³ The main new finding of Section 4 is that the entire rise in the between-firm component of the variance is due to a change in worker composition. This comes about equally from a rise in the variance of the average worker effect (35.6%) and the covariance of worker and firm effects (31.4%). In contrast, the worker-weighted variance of firm fixed effects does not rise, and at best declines early during our sample.

To reveal more about the rise in the covariance of worker and firm effects, the first two panels of Figure 7 display the joint distribution among deciles of worker and firm effects in 1980-1986 and 2007-2013, while the third panel shows the difference between them. The change in the pattern of sorting is striking. Over time, there has been a substantial shift of middle-decile individuals toward middle- and lower-decile firms, whereas the top-two-decile individuals have shifted toward top-decile firms. Hence, this increased sorting of workers has occurred across the entire firm and worker distribution. The figure also indicates there appears to have been a shift in employment away from high-wage firms. This partly helps to explain the decline in the worker-weighted variance of the firm fixed effect we found in Table 2.

Table 3 replicates these findings by firm size, which we discuss further in Section 5 below. First, we see in columns (1) to (6) that once we drop firms with 10,000 employees or more, the share of inequality accounted for by the between-firm component rises to 87.4% (row labeled “Between-Firm Variance”). Breaking this figure down, we see that about half of this (42.6%) comes from the increased dispersion of average individual effects, and the other half mostly comes from increased employee sorting across firms (33%), with some small additional contributions from the covariance of individual characteristics at the firm level (6.3% + 3.6% = 9.9% in total). If we drill down further into this group of firms in the right panel, keeping only firms with 1,000 employees or fewer (columns 7 to 12), we find the increase in inequality is entirely (102.1%) explained by a rise in the between-firm component. This comes about equally from two sources: the increased variance of the average worker effect (52.5%) and the increasing covariance of worker and firm effects (31.9%), plus small additional contributions from the firm effects (4.9%) and employee characteristics 7.1% + 5.2% = 12.3%.

²³This number is quite close to the corresponding statistic quoted in Subsection 3.1. However, statistics in this section may differ from those in Section 3.1 because of small differences in the sample selection and time periods analyzed.
Figure 7 – Distribution of Workers among Deciles of Worker and Firm Fixed Effects

(A) 1980-1986

(B) 2007-2013

(c) Change from 1980-1986 to 2007-2013

Notes: Calculations based on SSA data. See section 4.2 for sample selection criteria. Firm and worker fixed effects from our AKM estimation sorted into deciles. Since higher fixed-effect firms are larger, there are more employees in the higher firm fixed-effect deciles. Firm fixed effect deciles are computed with respect to the distribution of firms. Within each firm FE decile group, worker FE deciles are order from left to right from 1 to 10.
Finally, Table 3 also shows that larger firms experienced more substantial growth in inequality inside the firm (a change of 5.2 log points overall, but a reduction of 0.1 log points in firms with less than 1,000 employees). Given a similar absolute increase in between-firm inequality, this implies that larger firms experienced stronger increases in overall earnings inequality than smaller firms. (The same fact could also be seen by comparing the panels of Figure 2, which plot the entire time series during this period.) Interestingly, initially large firms appear to have had lower within-firm inequality than smaller firms. This is consistent with the view that large firms may have compressed wages, at least for the bottom end of their workforce. Yet, by the end of our sample period, there is no difference in earnings inequality between large and small firms (row 1).  

We also examined to what extent our main findings in Table 2 can be explained by employment shifts between industries. While there are some interesting differences in the time trends in the variance components across industries, our three main patterns of a rise in sorting, a rise in the variance of worker effects, and a stagnation (or small reduction) in the variance of firm effects occur within major sectors. Hence, most of our findings are driven by changes within sectors, and changes in sector composition have only a moderate effect.

Figure 8 shows the evolution of the correlation between worker and firm fixed effects by major industries for five seven-year time periods covering the period 1980 to 2013. All but three major industries see strong increases in correlations beginning in the early 1980s that are slowing down over time. The three exceptions are agriculture, public administration, and construction. Table 4 shows the corresponding values of the correlation, the covariance, and the variances of worker and firm effects, respectively. We also performed a simple counterfactual exercise that recalculated the variances and the correlation, holding constant the share of 1-digit and 4-digit industries (not shown). Secular sectoral employment shifts cannot explain any of the increase in the correlation of worker and firm effects at the aggregate level. The only component of the composition of the variance in earnings that is affected by industry shifts is the rise in the variance of the worker fixed effect, about one-third of which is explained by industry employment shifts. This component can explain the entire impact of sectoral shifts on the increase in the total variance of earnings that we find (not shown) and that has been documented elsewhere.

\[24\text{It is worth noting that even smaller firms experience an increase in the average within-firm variance of worker effects. However, this is largely offset by a reduction in the variance of the residual, and in the reduction in the covariance of worker effects and time-varying worker characteristics within firms.} \]
Explaining Trends in within- and between-Firm Inequality

In this section, we first discuss factors that are associated with between-firm inequality, chiefly a rise in sorting and segregation, and then turn to changes in within-firm inequality, especially within larger firms.

5.1 Accounting for between-Firm Inequality

An important question for understanding the strong rise in inequality in the United States is which mechanisms underlie a rise in assortative matching of high-wage workers to high-wage firms. Any hypothesis should also be compatible with: a stable distribution of $\psi^j$—firm differences in composition-adjusted pay; a rise in worker sorting and worker segregation; a relatively stable distribution of firm size (see Figure A.17—so this is not simply the atomization of firms; in fact, firm size has grown modestly during this period), and the fact that a rise in sorting and segregation is occurring within industries, regions, and demographic groups. Our finding in the statistical analysis (Section 4) that
the variance of the worker-firm (“match”) component is stable over time provides additional discipline on possible models. There are several candidate explanations, including outsourcing, changes in rent sharing, changes in search costs, or technological or organizational innovations that arise in worker-firm or worker-worker complementarities, but it is difficult to fit all of our facts within any basic model.

One explanation is that rising overall inequality is driven by skill-biased technical change, whereas rising outsourcing is constraining the impact on within-firm inequality. Likely drivers of the rise in outsourcing include falling costs of outsourcing (due to improving information-communications technology), a desire to limit the extent of inequality within firms due to concerns over fairness (e.g., Akerlof and Yellen (1990) and Weil (2014)), and a push by businesses to focus on “core competencies” (Prahalad and Hamel (1990)). This would lead firms to reorganize away from full-service production toward a more focused occupation structure. This is consistent with findings that occupations are increasingly concentrating within industries and firms (Kremer and Maskin (1996), Handwerker (2015)) and industries, as shown in Figure 9. The rise in outsourcing is also consistent with the increased occupational, educational, and ability segregation of employees found in Sweden by Håkanson et al. (2015), in Germany by Card et al. (2013), and in the United States by Barth et al. (2014). Goldschmidt and Schmieder (2015) examine German data, finding clear evidence that a rise in outsourcing contributed to increasing inequality. An explanation based on outsourcing could also be compatible with a stable distribution of firm fixed effects and firm size, especially in the United States, where existing low-wage firms could absorb outsourced workers.

Of course, other possible stories can also generate sorting and segregation. One class of models posits that firms pay different wages to workers with the same skills. This idea has a long tradition in labor economics, going back at least to Slichter (1950) and Stigler (1961), and lies at the heart of modern search theory. Several explanations have been put forward for such wage differentials, including that firms share product market rents, the presence of monopsony power in the labor market, or the presence of efficiency wage setting. If high-wage workers are more mobile or have a higher elasticity of labor supply, they will be more likely to work at high-wage employers (e.g., Card et al. (2016)). At a given distribution of firm pay, a rise in relative search efficiency or elasticities of high-wage with respect to low-wage workers will lead to an increase in sorting.

---

25While the concept of “core competencies” may not be well known in economics, this is an extremely popular idea in the business and consulting world; the Prahalad and Hamel (1990) article that coined the term has received almost 30,000 citations as of October 2016.
Notes: This figure plots the median Herfindahl-Hirschman concentration index (HHI) of occupations by industry in the CPS. Because of changes in the occupational classification system in 1982, 2002, and 2012, the figure is spliced across these three years and is normalized to zero in 1981 and 1982. Only individuals aged 20-60; who earn a positive wage income in the given year; who work at least 35 hours per week for 40 weeks; and who are not in education, public administration, or military industries are included.

An alternative class of explanations of the presence of assortative matching rests on complementarities in production between high-wage workers and high-wage firms. In a competitive market, such an interaction would lead to a wage premium for high-skilled workers at high-wage firms. This is consistent with our finding in Section 4 that worker-firm interactions contribute to explaining the variance of wages. A growing number of papers suggest that search frictions may prevent the labor market from reaching its optimal allocation of workers to firms. A reduction in search costs, perhaps due a decline in the cost of acquiring information or a rise in labor market intermediation, could then raise the degree of assortative matching. At a given type of technology, the increase in the effective supply of high-wage workers would tend to depress the match component, consistent with the stability of its variance we found in Section 4.
A related set of explanations rely on the interaction of fixed technologies with a rising dispersion of worker productivity. Kremer and Maskin (1996) suggest that complementarities in production occurred between low- and high-skilled workers. A similar version of this explanation posits that firms need to limit within-firm inequality for reasons of fairness or because of benefit-related fixed costs. At a given complementarity, a rise in the relative efficiency (or relative supplies) of high- and low-skilled workers implies that firms will hire increasingly homogeneous workers, leading to a pattern of segregation. If firms with higher complementarity also pay higher average wages, such a mechanism could explain an increase in sorting. A model by Acemoglu (1999) features search frictions, and in this model, firms decide what type of job to open before meeting a worker. When worker skills are similar, a pooling equilibrium emerges in which firms create “middling” jobs. With higher skill dispersion, it becomes optimal for firms to open good and bad jobs, suitable for high- and low-skill workers, respectively, leading to a separating equilibrium.

A closely related class of explanations relies on a process of uneven adoption of technological innovations between firms. Insofar as innovating firms require higher-skilled workers, this would imply a rise in the worker-firm complementarity for some firms and could explain an increase in either sorting or segregation. Whether such uneven technological progress would also imply a change in firm pay differentials or a change in worker-firm components in pay depends on the particular wage setting mechanism. While innovating firms may raise wages to attract high-skilled workers, they may also differentiate themselves or the jobs they offer to high-skilled workers in terms of their benefits. Free food, drinks, and massages at Google or tree houses at Amazon come to mind, as do companies like Starbucks that offer free college tuition and free food (and of course free coffee).

5.2 Accounting for within-Firm Inequality

To investigate within-firm inequality, we focus on larger firms, because as noted in Sections 3.4 and 4.3, the within-firm increases in inequality have primarily occurred in large firms. To examine why these mega (10k+ employee) firms see this much greater increase in inequality, in Figure 10 we plot the change in earnings for employees in various positions in the firm, ranging from the top-paid employee down to the median-paid employee. We see two clear differences between larger and smaller firms (here, firms with between 100 and 1,000 employees). First, large firms saw a fall in real median earnings of 0.07 (7%), while smaller firms saw an increase of 0.27 (31%). Second, in
larger firms, pay increases at the top end were far larger: the highest-paid employees in larger firms saw their average log pay increase by 86 log points (137%) in real terms since 1981, while the top-paid employees in smaller firms saw an increase of 0.37 (45%). Hence, the gap in real pay increases between the median and top employees in the mega-firms between 1981 and 2013 was 94 log points (155%) compared with 10 log points (11%) in smaller firms, a strikingly large difference. We now turn to examining these two changes in large firms.

5.2.1 Stagnating Earnings for Lower-Paid Workers in Large Firms

Figure 10 showed that in firms with 10,000+ workers, median pay has fallen by 7% in real terms between 1981 and 2013 (compared with a rise of 31% in firms with 100 to 1,000 employees). This collapse in earnings in the bottom 50% of large firms is an important force driving rising within-firm inequality in these firms because of the large share of these employees. The question is, why has pay fallen more in the lower percentiles of large firms compared with small firms?26

One fact that helps to explain this inequality is that the lower percentile earnings in large firms have converged from above with those in smaller firms. So, for example, in 1981 the median-paid employee in 10,000+ employee firms was paid 40 log points more than the median paid employee in firms with 100 to 1,000 employees, but this gap has shrunk to 5 log points by 2013.

To examine this convergence in pay for lower-earning employees in large firms, we used the CPS, which has had information on firm size since 1987. As shown in Figure 11, we find that the earnings premium for low-skilled employees (high school or less) in large firms (1000+ employees using the CPS definition) compared with small firms (fewer than 100 employees) has fallen by over half, from 36% in 1987 to 15% in 2013. In comparison, higher-skilled employees (the rest of the population, who have at least some college education) have seen this earnings premium fall far less, from 29% in 1987 to 23% in 2013. Hence, the earnings premium for low-skilled employees in large firms has fallen by 21 percentage points in the last 27 years (1987-2013), potentially accounting for much of the 35% difference in growth in median earnings between the very largest firms and the rest over the last 33 years (1981-2013) seen in the SSA data.
Figure 10 – Change in within-Firm Distribution of Annual Earnings: Smaller and Larger Firms

(A) Workers at Firms with 100 to 1,000 employees

(B) Workers at Firms with more than 10,000 employees

Notes: Only firms and individuals in firms of the listed size are included. Only full-time individuals aged 20 to 60 are included in all statistics, where full-time is defined as earning the equivalent of minimum wage for 40 hours per week in 13 weeks. Individuals and firms in public administration or educational services are not included. Statistics shown are based on the average log earnings among those at the given rank or percentile within their firm. All values are adjusted for inflation using the PCE price index.
Figure 11 – The Pay Premium in Larger Firms, by Education

(A) High school or less

(B) At least some college

Notes: Data are from the Current Population Survey Annual Social and Economic Supplement. Only individuals aged 20-60; who earn a positive wage income in the given year; who work at least 35 hours per week for 40 weeks; and who are not in education, public administration, or military industries are included. High school or less refers to those who have no more education than a high school diploma or equivalent. At least some college refers to the remainder of the population: those with at least one year of college education. Values shown are the differences in mean log earnings among those in the given firm size bracket, compared with those in firms with fewer than 100 workers.

5.2.2 Rising Earnings in the Top 1%

The other striking exception from the between-firm inequality result in Section 3 was the large gap between the earnings growth of the top 1% and the rest of the firm, particularly among the top 0.5% paid employees. This top 0.5% rising pay phenomenon was particularly striking in the largest (10,000+ employee) firms. To help explain this result, Figure 12 plots the coefficients from regressions of the yearly change in log earnings for top earners at different positions in firms of different sizes on the annual returns on the S&P 500 plus controls for GDP growth and unemployment. For example, the top right point with a triangle marker on the yellow “10k+” line indicates that the highest-paid employees in firms with 10,000 or more employees saw their annual log earnings change in relation to S&P 500 returns with a coefficient of 0.38. That is, for every 10% the S&P 500 rose, their earnings rose by 3.8%. Figure 12 shows how the earnings of the highest-paid employees at the 10,000+ employee firms have very high coefficients: 0.38 for the highest-paid employee (presumably the CEO), 0.3 for the second highest-paid employee (presumably the CFO), down to 0.15 for the 50th highest-paid employee (a very senior

---

26This is related to the general debate over the collapsing large firm wage premium (see, e.g., Cobb et al. (2016))
manager). In comparison, top-paid employees in firms of 100 to 1,000 employees saw a compensation connection with the returns on the S&P 500 of about 0.08.

One possible explanation for these results is that 10,000+ employee firms often reward their senior executives with stock options and stock grants. Moreover, this stock-based remuneration (which is included in the W-2 earnings figure as long as options are exercised) has been rising over time. For example, in 2014 the annual compensation of the top-five executives listed in the Execucomp database—which spans roughly the top 1,800 largest U.S. firms by market capitalization—was 48% from stock options and stock awards, up from 15% in 1993 (the first full year of Execucomp data). Alongside this rising stock payment to senior executives, there has been a 19-year stock market bull run, with real returns averaging 9.5% between 1981 and 1999. As Figure 10 shows, the senior executives at the largest firms received extremely generous pay increases over this 1981-1999 period, and since 2000 (a period of low stock returns) increases have moved roughly in line with the rest of their firm. Thus, it appears that the top 50 or so executives in the largest U.S. firms have experienced rapidly rising earnings—far outstripping

**Figure 12 – Earnings responsiveness to the S&P 500 returns**

Notes: Only full-time individuals aged 20 to 60 are included in all statistics, where full-time is defined as earning the equivalent of minimum wage for 40 hours per week in 13 weeks. Individuals and firms in public administration or educational services are not included. Each data point represents a regression coefficient; the dependent variable for each regression is the change in average log earnings from year \( t \) to \( t+1 \) among those at the given rank or percentile within their firm, for firms of given sizes. The coefficient shown is on the log change in the S&P 500 during year \( t \). There are 35 observations in each regression: one per year from 1979 to 2013. The regression includes controls for unemployment in year \( t \) and log GDP growth between year \( t \) and year \( t+1 \). All values are adjusted for inflation using the PCE price index.
their colleagues—in part because of the rising level and generosity of stock-based compensation.\footnote{Simply applying the magnitudes of the 680\% real increase in the S&P 500 over the period 1981–1999 to the average 0.25 coefficient on the S&P 500 returns in Figure 12 yields a real cumulative pay increase of 170\%, which is similar to the earnings gains of up to 200\% that this group made over the same period (see Figure 10). One outstanding question this analysis raises, however, is why these stock-driven pay rises have been permanent, rather than one-off high earnings payouts during the years of unexpectedly strong S&P 500 performance. One recent paper offering an explanation is \textcite{Shue2016}, who report that S&P 500 firms tend to give executives similar numbers of stock options each year, despite these options rising in value with the firms’ stock price. Hence, historic rises in the S&P 500 tend to get locked into future equity pay levels.}

Given this, it is perhaps surprising that in terms of contributions to rising inequality, the top 50 employees account for only 3\% of the increase in inequality in mega-firms. The reason is they are a small share of employment—accounting for only 20,000 of the 20 million employees in mega-firms—so have little impact on the increasing within firm variance of log wages. However, the top-earning 10\% of employees, a group that contains a much wider group of managers, technicians, and other highly paid individuals in large firms, accounts for 45\% of increasing within-firm inequality.

6 Conclusions

Using a massive, new, matched employer-employee database that we construct for the United States, we documented three stylized facts. First, the rise in earnings inequality between workers over the last three decades has primarily been a between-firm phenomenon. Two-thirds of the increase in the variance of log earnings from 1981 to 2013 can be accounted for by differences in earnings between firms and only one-third by differences between workers within firms.

Second, examining the sources of this increase in between-firm inequality, we find that it has been driven about equally by increased employee sorting (i.e., high-wage workers are increasingly found at high-wage firms) and segregation (i.e., highly paid employees are increasingly clustering in high-wage firms with other high-paid workers, while low-paid employees are clustering in other firms). These two phenomena also seem to be happening globally, with similar patterns seen in every country for which detailed worker-firm earnings data are available (i.e., Brazil, Germany, Sweden, Japan and the United Kingdom). Third, the rise in within-firm inequality is concentrated in large firms with 1000+ (and particularly 10,000+) employees. This is driven both by a fall in the pay premium in large firms for median- and lower-paid employees, and also by rising pay for the top 10\% of employees.
These results raise the question as to what is driving this dramatic change in worker composition across firms. While our analysis does not provide a definitive answer to this question, a variety of circumstantial evidence indicates that outsourcing could be playing an important role in allowing firms to constrain inequality within firms and focus on core competency activities, spinning off nonessential activities such as cleaning, catering, security, accounting, IT, and HR. Since firm size is only slowly growing over this period, firms are not atomizing; instead they may be reorganizing around a more concentrated set of occupations, perhaps leading to greater cross-firm segregation by worker skill level. Studying this and other channels is an important area for further research.

Finally, this increase in between-firm inequality raises a question over its impact on individual welfare. We believe increased worker sorting and worker segregation is potentially worrisome for several reasons. One concern is of course that low-wage workers appear to have lost access to good jobs at high-wage firms, increasing overall aggregate inequality. Another concern is that firms play an important role in providing employee health care and pensions, so rising earnings segregation could very well spill into rising health care and retirement inequality. Indeed, over the last 30 years, as noted by the National Academies of Sciences, Engineering, and Medicine (2015), the correlation between income and life expectancy has increased greatly at the same time as a greater fraction of wealth for those at the top comes from benefits, including health insurance. Moreover, given the importance of work experience to long-run earnings growth, if employees gain experience more rapidly by working alongside higher-ability colleagues, then rising segregation will dynamically increase inequality.
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Variance Component</td>
<td>Share of Total</td>
<td>Variance Component</td>
<td>Share of Total</td>
<td>Variance Component</td>
<td>Share of Total</td>
</tr>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
<td>(4)</td>
<td>(5)</td>
<td>(6)</td>
</tr>
<tr>
<td></td>
<td>Var(log(y))</td>
<td>0.748</td>
<td>100</td>
<td>0.855</td>
<td>100</td>
<td>0.861</td>
</tr>
<tr>
<td>Between-Firm</td>
<td>Var(m_we_f)</td>
<td>0.083</td>
<td>11.1</td>
<td>0.106</td>
<td>12.4</td>
<td>0.121</td>
</tr>
<tr>
<td>Components</td>
<td>Var(firm effect)</td>
<td>0.120</td>
<td>16.0</td>
<td>0.115</td>
<td>13.5</td>
<td>0.099</td>
</tr>
<tr>
<td>of Variance</td>
<td>Var(m_xb_f)</td>
<td>0.009</td>
<td>1.2</td>
<td>0.013</td>
<td>1.5</td>
<td>0.012</td>
</tr>
<tr>
<td></td>
<td>2cov(m_we_f,firm)</td>
<td>0.012</td>
<td>1.6</td>
<td>0.035</td>
<td>4.1</td>
<td>0.049</td>
</tr>
<tr>
<td></td>
<td>2cov(m_we_f,m_xb_f)</td>
<td>0.015</td>
<td>2.0</td>
<td>0.022</td>
<td>2.5</td>
<td>0.021</td>
</tr>
<tr>
<td></td>
<td>2cov(firm,m_xb_f)</td>
<td>0.020</td>
<td>2.7</td>
<td>0.025</td>
<td>2.9</td>
<td>0.021</td>
</tr>
<tr>
<td>Within-Firm</td>
<td>Var(diff_we_f)</td>
<td>0.272</td>
<td>36.4</td>
<td>0.309</td>
<td>36.2</td>
<td>0.325</td>
</tr>
<tr>
<td>Components</td>
<td>Var(diff_xb_f)</td>
<td>0.048</td>
<td>6.4</td>
<td>0.065</td>
<td>7.6</td>
<td>0.076</td>
</tr>
<tr>
<td>of Variance</td>
<td>Var(diff_r_f)</td>
<td>0.151</td>
<td>20.3</td>
<td>0.158</td>
<td>18.5</td>
<td>0.149</td>
</tr>
<tr>
<td></td>
<td>2cov(diff_we_f,diff_xb_f)</td>
<td>0.017</td>
<td>2.2</td>
<td>0.009</td>
<td>1.0</td>
<td>-0.011</td>
</tr>
<tr>
<td>Raw Decomposition</td>
<td>Between Firm Var</td>
<td>0.488</td>
<td>65.3</td>
<td>0.540</td>
<td>63.2</td>
<td>0.538</td>
</tr>
<tr>
<td></td>
<td>Within Firm Var</td>
<td>0.488</td>
<td>65.3</td>
<td>0.540</td>
<td>63.2</td>
<td>0.538</td>
</tr>
<tr>
<td>Segregation Index</td>
<td>Var(m_we_f)/Var(we)</td>
<td>0.235</td>
<td>0.255</td>
<td>0.271</td>
<td>0.273</td>
<td>0.273</td>
</tr>
</tbody>
</table>

Notes.
- log(y) - natural log of annual earnings
- m_we_f - mean worker effect across firms
- m_xb_f - mean Xb across firms
- m_r_f - mean residual across firms
- diff_we_f - difference of worker effect from mean worker effect across firms
- diff_xb_f - difference of Xb from mean x across firms
- diff_r_f - difference of the residual from the mean residual across firms

Estimates are from the baseline 100% sample - 520 hrs at minimum wage, male, age 20-60, with normalization: age'=(age-40)/40.

Raw decomposition refers to the between and within firm variance composition simply on log wages, rather than using the CHK components.
Table 3: Decomposition of the Rise in Earnings Inequality Between- and Within-Firms by Firm Size

<table>
<thead>
<tr>
<th></th>
<th>Interval 1 (1980-1986)</th>
<th>Interval 5 (2007-2013)</th>
<th>Change from 1 to 5</th>
<th>Sample Excluding Firms with Greater than 10,000 Employees</th>
<th>Sample Excluding Firms with Greater than 1,000 Employees</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Variance Component</td>
<td>Share of Total</td>
<td>Variance Component</td>
<td>Share of Total</td>
<td>Variance Component</td>
</tr>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
<td>(4)</td>
<td>(5)</td>
</tr>
<tr>
<td>Var(log(y))</td>
<td>0.774</td>
<td>100</td>
<td>0.912</td>
<td>100</td>
<td>0.138</td>
</tr>
<tr>
<td>Between-Firm</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Var(m_we_f)</td>
<td>0.100</td>
<td>12.9</td>
<td>0.159</td>
<td>17.4</td>
<td>0.059</td>
</tr>
<tr>
<td>Components of Variance</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Var(firm effect)</td>
<td>0.114</td>
<td>14.7</td>
<td>0.116</td>
<td>12.7</td>
<td>0.002</td>
</tr>
<tr>
<td>2cov(m_we_f,firm)</td>
<td>0.007</td>
<td>0.9</td>
<td>0.053</td>
<td>5.8</td>
<td>0.046</td>
</tr>
<tr>
<td>2cov(m_we_f,m_xb_f)</td>
<td>0.013</td>
<td>1.7</td>
<td>0.022</td>
<td>2.4</td>
<td>0.009</td>
</tr>
<tr>
<td>2cov(firm,m_xb_f)</td>
<td>0.017</td>
<td>2.2</td>
<td>0.022</td>
<td>2.4</td>
<td>0.005</td>
</tr>
<tr>
<td>Within-Firm</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Var(diff_we_f)</td>
<td>0.291</td>
<td>37.6</td>
<td>0.339</td>
<td>37.2</td>
<td>0.048</td>
</tr>
<tr>
<td>Components of Variance</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Var(diff_xb_f)</td>
<td>0.049</td>
<td>6.3</td>
<td>0.053</td>
<td>5.8</td>
<td>0.004</td>
</tr>
<tr>
<td>2cov(diff_we_f,diff_xb_f)</td>
<td>0.013</td>
<td>1.7</td>
<td>0.002</td>
<td>0.2</td>
<td>-0.012</td>
</tr>
<tr>
<td>Within Firm Variance</td>
<td>0.514</td>
<td>66.4</td>
<td>0.532</td>
<td>58.4</td>
<td>0.018</td>
</tr>
<tr>
<td>Raw Decomposition</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Between Firm Var</td>
<td>0.259</td>
<td>33.5</td>
<td>0.381</td>
<td>41.7</td>
<td>0.121</td>
</tr>
<tr>
<td>Within Firm Var</td>
<td>0.514</td>
<td>66.5</td>
<td>0.531</td>
<td>58.3</td>
<td>0.017</td>
</tr>
<tr>
<td>Segregation Index</td>
<td>Var(m_we_f)/Var(we)</td>
<td>0.255</td>
<td>0.319</td>
<td>0.064</td>
<td>0.285</td>
</tr>
</tbody>
</table>

Notes.

- log(y) - natural log of annual earnings
- m_we_f - mean worker effect across firms
- m_xb_f - mean Xb across firms
- m_r_f - mean residual across firms
- diff_we_f - difference of worker effect from mean worker effect across firms
- diff_xb_f - difference of Xb from mean Xb across firms
- diff_r_f - difference of the residual from the mean residual across firms

Estimates are from the baseline 100% sample - 520 hrs at minimum wage, male, age 20-60, with normalization: age'=((age-40)/40).

Raw decomposition refers to the between and within firm variance composition simply on log wages, rather than using the CHK components.
<table>
<thead>
<tr>
<th>Industry</th>
<th>Interval 1: 1980-1986</th>
<th></th>
<th></th>
<th>Change from Interval 1 to 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Agriculture, Forestry, Fishing</td>
<td>2.12</td>
<td>0.29</td>
<td>0.12</td>
<td>-0.04</td>
</tr>
<tr>
<td>Mining</td>
<td>1.69</td>
<td>0.32</td>
<td>0.10</td>
<td>0.01</td>
</tr>
<tr>
<td>Construction</td>
<td>7.82</td>
<td>0.41</td>
<td>0.10</td>
<td>0.01</td>
</tr>
<tr>
<td>Manufacturing</td>
<td>25.83</td>
<td>0.30</td>
<td>0.08</td>
<td>0.04</td>
</tr>
<tr>
<td>Transportation &amp; Public Utilities</td>
<td>7.94</td>
<td>0.27</td>
<td>0.09</td>
<td>0.07</td>
</tr>
<tr>
<td>Wholesale Trade</td>
<td>5.92</td>
<td>0.40</td>
<td>0.08</td>
<td>-0.06</td>
</tr>
<tr>
<td>Retail Trade</td>
<td>10.25</td>
<td>0.41</td>
<td>0.09</td>
<td>-0.06</td>
</tr>
<tr>
<td>Finance, Insurance, Real Estate</td>
<td>4.95</td>
<td>0.47</td>
<td>0.12</td>
<td>0.02</td>
</tr>
<tr>
<td>Services</td>
<td>16.05</td>
<td>0.44</td>
<td>0.14</td>
<td>0.05</td>
</tr>
<tr>
<td>Public Administration</td>
<td>7.65</td>
<td>0.22</td>
<td>0.08</td>
<td>0.05</td>
</tr>
<tr>
<td>Unknown</td>
<td>9.77</td>
<td>0.33</td>
<td>0.10</td>
<td>-0.04</td>
</tr>
</tbody>
</table>

Notes: Variance and correlation of fixed effects estimated by AKM model as explained in Section 4.
Table C1: Summary Statistics for Overall Sample and Individuals in Largest Connected Set

<table>
<thead>
<tr>
<th>7 Year Interval</th>
<th>All full-time men, age 20-60</th>
<th>Individuals in largest connected set</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Number worker/yr. obs. (1)</td>
<td>Number workers (2)</td>
</tr>
<tr>
<td>1980-1986</td>
<td>336,095,166</td>
<td>66,253,680</td>
</tr>
<tr>
<td>Ratio: largest connected/all</td>
<td>99.0</td>
<td>98.6</td>
</tr>
<tr>
<td>1987-1993</td>
<td>378,061,870</td>
<td>71,852,732</td>
</tr>
<tr>
<td>Ratio: largest connected/all</td>
<td>98.9</td>
<td>98.5</td>
</tr>
<tr>
<td>Between Firm Variance</td>
<td>98.9</td>
<td>98.5</td>
</tr>
<tr>
<td>1994-2000</td>
<td>408,883,618</td>
<td>76,590,263</td>
</tr>
<tr>
<td>Ratio: largest connected/all</td>
<td>98.7</td>
<td>98.4</td>
</tr>
<tr>
<td>2001-2006</td>
<td>431,749,460</td>
<td>81,950,386</td>
</tr>
<tr>
<td>Ratio: largest connected/all</td>
<td>98.4</td>
<td>98.1</td>
</tr>
<tr>
<td>2007-2013</td>
<td>422,334,068</td>
<td>82,761,555</td>
</tr>
<tr>
<td>Ratio: largest connected/all</td>
<td>98.1</td>
<td>97.8</td>
</tr>
<tr>
<td>Change from first to last interval</td>
<td>0.106</td>
<td>0.061</td>
</tr>
</tbody>
</table>

Notes:
A 100% sample of the SSA Master Earnings File. Male only, ages 20-60, over 520 hours worked at minimum wage.
The dependent variable is log annual earnings at main job.
Annual earnings are adjusted for inflation with the PCE deflator at base year 2013.
### Table C2: Estimation Results for AKM Model, Fit by Interval

<table>
<thead>
<tr>
<th>Sample</th>
<th># Worker Effects (1)</th>
<th># Worker Effects (2)</th>
<th># Worker Effects (3)</th>
<th># Worker Effects (4)</th>
<th># Worker Effects (5)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>65,311,438</td>
<td>70,802,234</td>
<td>75,365,461</td>
<td>80,409,315</td>
<td>80,920,372</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Summary</th>
<th># Firm Effects (1)</th>
<th># Firm Effects (2)</th>
<th># Firm Effects (3)</th>
<th># Firm Effects (4)</th>
<th># Firm Effects (5)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>5,206,888</td>
<td>5,733,422</td>
<td>5,884,023</td>
<td>5,889,906</td>
<td>5,258,436</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Statistics</th>
<th>Sample size (1)</th>
<th>Sample size (2)</th>
<th>Sample size (3)</th>
<th>Sample size (4)</th>
<th>Sample size (5)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>332,624,208</td>
<td>373,806,862</td>
<td>403,724,055</td>
<td>425,006,770</td>
<td>414,466,857</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Statistics</th>
<th>SD(log(y)) (1)</th>
<th>SD(log(y)) (2)</th>
<th>SD(log(y)) (3)</th>
<th>SD(log(y)) (4)</th>
<th>SD(log(y)) (5)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.865</td>
<td>0.925</td>
<td>0.928</td>
<td>0.960</td>
<td>0.956</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Summary of AKM Parameter</th>
<th>SD(WE) (1)</th>
<th>SD(WE) (2)</th>
<th>SD(WE) (3)</th>
<th>SD(WE) (4)</th>
<th>SD(WE) (5)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.596</td>
<td>0.644</td>
<td>0.667</td>
<td>0.692</td>
<td>0.695</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Estimates</th>
<th>SD(FE) (1)</th>
<th>SD(FE) (2)</th>
<th>SD(FE) (3)</th>
<th>SD(FE) (4)</th>
<th>SD(FE) (5)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.346</td>
<td>0.339</td>
<td>0.314</td>
<td>0.332</td>
<td>0.329</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Estimates</th>
<th>SD(Xb) (1)</th>
<th>SD(Xb) (2)</th>
<th>SD(Xb) (3)</th>
<th>SD(Xb) (4)</th>
<th>SD(Xb) (5)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.238</td>
<td>0.278</td>
<td>0.297</td>
<td>0.254</td>
<td>0.256</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Estimates</th>
<th>Between Firm Variance (1)</th>
<th>Between Firm Variance (2)</th>
<th>Between Firm Variance (3)</th>
<th>Between Firm Variance (4)</th>
<th>Between Firm Variance (5)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.030</td>
<td>0.080</td>
<td>0.116</td>
<td>0.131</td>
<td>0.141</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.113</td>
<td>0.084</td>
<td>0.027</td>
<td>0.084</td>
<td>0.091</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Estimates</th>
<th>Corr(FE,Xb) (1)</th>
<th>Corr(FE,Xb) (2)</th>
<th>Corr(FE,Xb) (3)</th>
<th>Corr(FE,Xb) (4)</th>
<th>Corr(FE,Xb) (5)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.124</td>
<td>0.133</td>
<td>0.115</td>
<td>0.129</td>
<td>0.143</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Estimates</th>
<th>RMSE(residual) (1)</th>
<th>RMSE(residual) (2)</th>
<th>RMSE(residual) (3)</th>
<th>RMSE(residual) (4)</th>
<th>RMSE(residual) (5)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.438</td>
<td>0.445</td>
<td>0.431</td>
<td>0.442</td>
<td>0.415</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Estimates</th>
<th>Adj R² (1)</th>
<th>Adj R² (2)</th>
<th>Adj R² (3)</th>
<th>Adj R² (4)</th>
<th>Adj R² (5)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.743</td>
<td>0.768</td>
<td>0.784</td>
<td>0.788</td>
<td>0.812</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Comparison</th>
<th>RMSE(match residual) (1)</th>
<th>RMSE(match residual) (2)</th>
<th>RMSE(match residual) (3)</th>
<th>RMSE(match residual) (4)</th>
<th>RMSE(match residual) (5)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.369</td>
<td>0.374</td>
<td>0.360</td>
<td>0.369</td>
<td>0.347</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Comparison</th>
<th>Adj R² (match residual) (1)</th>
<th>Adj R² (match residual) (2)</th>
<th>Adj R² (match residual) (3)</th>
<th>Adj R² (match residual) (4)</th>
<th>Adj R² (match residual) (5)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.817</td>
<td>0.836</td>
<td>0.850</td>
<td>0.852</td>
<td>0.868</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Comparison</th>
<th>SD(match effect) (1)</th>
<th>SD(match effect) (2)</th>
<th>SD(match effect) (3)</th>
<th>SD(match effect) (4)</th>
<th>SD(match effect) (5)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.260</td>
<td>0.266</td>
<td>0.264</td>
<td>0.267</td>
<td>0.244</td>
</tr>
</tbody>
</table>

Notes:
- A 100% sample of the SSA Master Earnings File. Male only, ages 20-60, over 520 hours worked at minimum wage.
- Age is normalized as age=(age-40)/40 as in CHK.
- We include an unrestricted set of year dummies and quadratic and cubic terms in age.
- We do not interact 5 education dummies with other covariates as in CHK.
- The dependent variable is log annual earnings at main job.
- Annual earnings are adjusted for inflation with the PCE deflator at base year 2013.
### Table C3: Basic Decomposition of the Rise in Inequality of Annual Earnings

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Variance Component</td>
<td>Share of Total</td>
<td>Variance Component</td>
<td>Share of Total</td>
<td>Variance Component</td>
<td>Share of Total</td>
</tr>
<tr>
<td><strong>Total Variance</strong></td>
<td>Var(log(y))</td>
<td>0.748</td>
<td>100</td>
<td>Var(WE)</td>
<td>0.355</td>
<td>47.5</td>
</tr>
<tr>
<td><strong>Components of Variance</strong></td>
<td>Var(WE)</td>
<td>0.355</td>
<td>47.5</td>
<td>Var(FE)</td>
<td>0.120</td>
<td>16.0</td>
</tr>
<tr>
<td></td>
<td>Var(Xb)</td>
<td>0.057</td>
<td>7.6</td>
<td>Var(residual)</td>
<td>0.151</td>
<td>20.3</td>
</tr>
<tr>
<td></td>
<td>2*Cov(WE,FE)</td>
<td>0.012</td>
<td>1.6</td>
<td>2*Cov(WE,Xb)</td>
<td>0.032</td>
<td>4.3</td>
</tr>
<tr>
<td></td>
<td>2*Cov(FE,Xb)</td>
<td>0.020</td>
<td>2.7</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Counterfactuals</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1. No rise in Corr(WE,FE)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2. No fall in Var(FE)</td>
<td>0.748</td>
<td></td>
<td></td>
<td>0.748</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3. Both 1 and 2</td>
<td>0.748</td>
<td></td>
<td></td>
<td>0.748</td>
<td></td>
</tr>
</tbody>
</table>

**Notes:**
Var(log(y)) - variance of annual earnings, Var(WE) - variance of worker fixed effects, Var(FE) - variance of firm fixed effects, Var(Xb) - variance of covariates.
See notes to Table 3 and C2.
References


Holzer, Harry J., Julia I. Lane, David B. Rosenblum, and Fredrik Andersson, Where are all the good jobs going? What national and local job quality and dynamics mean for US workers, Russell Sage Foundation, 2011.


Supplemental Online Appendix
Appendix: Additional Figures

A.1 Sensitivity and Robustness

Figure A.1 – Comparing the SSA totals to other records

(A) Total earnings

(B) Total employment

(C) Total firms

Notes: SSA data includes all entries in the MEF. National Income and Product Accounts (NIPA) data is from the St. Louis Federal Reserve Bank’s FRED service, series A576RC1, “Compensation of Employees, Received: Wage and Salary Disbursements.” Current Population Survey (CPS) total employment shows the yearly average of the monthly employment numbers in the CPS. This data is from the Bureau of Labor Statistics Table LNS12000000. Census firms shows the total number of firms reported by the Census Bureau’s Statistics of U.S. Businesses data set, available at http://www.census.gov/econ/susb/historical_data.html. All data are adjusted for inflation using the PCE price index.
**Figure A.2** – Comparing earnings variance in SSA and CPS data

(A) Levels

(b) Change since 1981

Notes: Only full-time individuals aged 20 to 60 are included in all statistics, where full-time is defined as earning the equivalent of minimum wage for 40 hours per week in 13 weeks. Individuals and firms in public administration or educational services are not included. Only firms and individuals in firms with at least 20 employees are included in SSA data.

**Figure A.3** – Cumulative distribution of annual earnings in CPS data

Notes: For each percentile, statistics are based on the minimum earnings among individuals in that percentile of earnings in each year. All values are adjusted for inflation using the PCE price index. Only full-time individuals aged 20 to 60 are included in all statistics, where full-time is defined as earning the equivalent of minimum wage for 40 hours per week in 13 weeks. Individuals and firms in public administration or educational services are not included.
Figure A.4 – Robustness: Controlling for Geography at County Level

Notes: See notes for Figure 3. Data for a graph similar to 3 is calculated for each county (of the firm) in each year, then averaged together by year, weighting counties by employment.

A.2 Other figures

B Appendix: Data Procedures

B.1 Social Security Administration Data

As noted in Section 2, this paper uses data from SSA’s MEF database. We begin with an extract from this file that includes one observation for each year, for each individual, for each firm that this individual worked for. (For self-employed individuals, the data set also contains these earnings from the IRS as reported in Schedule-SE tax form by the individuals. Because our focus is on firms with employees, we exclude these earnings from our analysis.) For each observation, this file includes the year, a transformation of that individual’s Social Security Number, along with the associated sex and date of birth; and the EIN, along with the associated 4-digit SIC code and state.

The first step we take with this data is to exclude individuals who did not have a reasonably strong labor market attachment in a given year from the analysis for that year. More
Figure A.5 – Robustness by region

(A) Northeast

(B) Midwest

(c) South

(d) West

Notes: See notes for Figure 3. Regions are based on Census region definitions. Percentiles are based on only those individuals in the given region.

Concretely, we consider an individual to be full-time in a given year and include in the analysis if, summing across all jobs, he/she earns at least the equivalent of 40 hours per week for 13 weeks at that year’s minimum wage (so $3,770 in 2013). (As discussed above, we also conducted robustness checks with other threshold levels, which show similar results; see Figure A.14.) This condition ensures that we are focusing on data about individuals with a reasonably strong labor market attachment, and that our results are comparable to other results in the wage inequality literature, such as Juhn et al. (1993) and Autor et al. (2008). The data from any individual earning below this threshold in a given year is excluded from all results for both firms and individuals in that year.

We assign workers to firms based on the firm where that worker earned the most money
Figure A.6 – Robustness to establishments versus firms - firms with 20+ workers

Notes: Source Census Longitudinal Business Database 1976-2013. All sectors excluding education and the public sector. Establishments only from firms with 20+ employees.

in a given year. Firm earnings statistics are based on total annual earnings of each individual whose primary job is with that firm, even if the worker earned part of that money in a different firm. Where our results analyze the same firm over multiple years, we include a correction to ensure that firms that change EINs are not counted as exiting in one year and entering in the next. We define an EIN in Year 1 as being the same firm as a different EIN in Year 2 if the following conditions are met. First, Year 1 must be the last year in which the original EIN appears, while Year 2 must be the first year that the new EIN appears in our data. Next, more than half of the individuals who worked in each firm must have also worked in the other firm. Finally, to ensure that our results aren’t influenced by a few individuals switching companies, we only include EINs in this switching analysis if they employ at least 10 individuals.

Firms are only included in our sample if they have at least 20 employees in a given year to ensure that firm-wide statistics are meaningful; for example, comparing an individual to the mean earnings at their two-person firm may not be a good way to characterize inequality within firms in a given year (though our results are robust to changing this threshold). We also exclude firms in the Educational Services (SIC Codes 8200 to 8299) and Public Administration (SIC Codes 9000 to 9899) industries, as employers in these industries are frequently not what we would consider firms. Finally, we exclude employers with EINs that begin with certain
two-digit codes that are associated with Section 218 Agreements, or other issues that may not be handled consistently in the data across years. Individuals whose primary job is with a firm in one of these excluded categories are also dropped from the data in that year.

In order to analyze a representative sample of individuals in a computationally feasible way, we analyze a one-eighth representative sample of all U.S. individuals from 1978 to 2013 (except in the firm and worker fixed effects analysis, in which we use a 100% sample). Results are robust to using a 100% sample; see Figure A.16a. The sample is organized as a longitudinal panel, in the sense that once an individual is selected into the sample, he/she remains in the sample until he/she dies. In particular, an individual is in our sample if the MD5 hash of a transformation of their Social Security Number begins with a zero or one; because MD5 hashes are hexadecimal numbers, this will select one in eight individuals. MD5 is a cryptographic algorithm that deterministically turns any string into a number that is essentially random. It is designed so that a slightly different input would lead to a completely different output in a way that is essentially impossible to predict. Because it took cryptographic researchers several years to figure out a way that, under certain circumstances, MD5 is somewhat predictable, this algorithm is certainly random enough for our purposes. Thus whether one individual is
Figure A.8 – Robustness by demographic (age group and gender)

(A) Age: 20 to 29

(B) Age: 30 to 39

(C) Age: 40 to 49

(D) Age: 50 to 60

(E) Men

(F) Women

Notes: See notes for Figure 3. Percentiles are based on all individuals, regardless of age or gender.
Figure A.9 – Robustness by industry: results by SIC 1-digit industry

(A) Agriculture, Mining, Construction & Other

(b) Manufacturing

(c) Utilities

(d) Trade

(E) Finance, insurance and real estate

(F) Services

Notes: See notes for Figure 3. Industries are based on definitions from https://www.osha.gov/pls/imis/sic_manual.html. Percentiles are based on only those individuals employed in the given industry.
**Figure A.10** – Robustness by industry: averaged across analysis within SIC 4-digit industry

Notes: See notes for Figure 3. Data for a graph similar to 3 is calculated for each 4-digit industry in each year, then averaged together by year, weighting industries by employment.

**Figure A.11** – Robustness by measure of firm average earnings

Notes: See notes for Figure 3. Individual statistics are the same for all lines; firm statistics are calculated differently, as indicated.
Figure A.12 – Continuing firms only

Notes: See notes for Figure 3. Only firms (and individuals in those firms) that are in the sample in both 1981 and 2013 are included in the analysis.

included in our sample is essentially independent of whether some other individual is included, regardless of how similar their SSNs are.

We top-code all variables of interest above the 99.999th percentile to avoid potential problems with disclosure or extreme outliers. Variables are top-coded with the average value (or geometric average value, as appropriate) of all observations within the top 0.001%. Variables are top-coded immediately before analysis. An exception is in analysis of top income ranks within firms, as in Figures 10 and 12, which could be more affected by top-coding; for these analyses, we top-code at the maximum value in Execucomp for the given year (or, before 1992, the average of the maximum values between 1992 and 1994). Top-coding at the 99.999th percentile has no visible effect on the main analysis: see Figure A.16b for results top-coding at the maximum value in Execucomp. Finally, we adjust all dollar values in the data set to be equivalent to 2013 dollars with the Personal Consumption Expenditure (PCE) price index.28

B.2 Current Population Survey Data

Micro data from the Current Population Survey (CPS) Annual Social and Economic Supplement, as made available by Flood et al. (2015). Data for year t is based on the survey from year t + 1. The sample is restricted to those aged between 20 and 60; with non-zero, non-missing wage and salary income; and who are not in education, public administration, or military industries. Figures in the text are restricted to those who had at least 35 usual hours of work per week and who worked at least 40 weeks. For comparability with SSA data, data for

28http://research.stlouisfed.org/fred2/series/PCEPI/downloaddata?cid=21
**Figure A.13 – Less-restrictive sample selection**

(A) All ages

(B) All firm sizes

(C) All industries

(D) All ages, firm sizes, and industries

Notes: See notes for Figure 3. Sample selection criteria are relaxed as indicated.

Figures A.2 and A.3 restrict to those earning at least the equivalent of 40 hours per week for 13 weeks at that year’s minimum wage. All statistics are weighted by the person-level supplemental weight. Wherever possible, we use variables that are coded consistently throughout the time period considered. Education data is based on variable EDUC; industry and occupation data are based on variables IND1990 and OCC1990, respectively.
Figure A.14 – Different definitions of full-time workers

(A) 6.5 weeks
(B) 26 weeks

(c) 52 weeks

Notes: See notes for Figure 3. Minimum earnings thresholds are adjusted to the equivalent of 40 hours per week at minimum wage for the given number of weeks.

C Appendix: The Abowd, Kramarz and Margolis decomposition

C.1 Identifying Assumption

Estimation of the firm effects in equation (3) crucially relies on earnings changes of workers switching employers. Hence, the estimated firm effects will capture any systematic differences in earnings of movers before and after the job move. This includes the difference in firm effects
Figure A.15 – Five year average earnings: comparing 1981-1985 vs 2009-2013

Notes: See notes for Figure 3. Only individuals who are in the sample for all five years are included. Individual income is calculated as the average of log earnings over the five years. Firm statistics are based on the average of mean log earnings at the firm that the individual was in, in each of the five years (even if that includes different firms).

Figure A.16 – Other restrictions

(A) 100% sample

(B) Winsorizing at Execucomp maximum

Notes: See notes for Figure 3.
between the sending and receiving firm, but also potential differences in average fixed worker-firm match effects, or systematic transitory earnings changes leading up to or following a job change. Hence, to associate estimated firm effects with true underlying firm-specific differences in pay, one has to assume that conditional on worker and firm effects, job moves do not depend systematically on other components. This assumption, often referred to as the conditional random mobility (CRM) assumption, and its relation to economic models of job mobility, is discussed at length in AKM and CHK, among others, and we will not review the theoretical arguments against or in favor here.

On a fundamental level, whether the CRM assumption is conceptually or empirically plausible or not, the estimation of the parameters in equation (3) is done by Ordinary Least Squares, and hence one relies on “random” variation provided by nature, not on known sources of manipulation. To ensure our core assumption and findings are plausible, following CHK, we will provide several pieces of corroborating evidence below. This includes event studies of the effect of worker mobility, the goodness of fit of the model, the value added of allowing for worker-firm match effects, and the properties of the residuals. After a careful review, we conclude from this evidence that there appear to be no large, systematic worker-firm or transitory components influencing job mobility. We thus join an increasing number of papers whose results indicate the AKM model can be estimated without systematic bias (e.g., AKM, CHK, Barth et al. (2014), Abowd et al. (2016)). Nevertheless, we are well aware of the limitations of the model, and incorporate it into our overall approach. Among other measurements, we will separately estimate worker-firm component in earnings $m_{ij}$, and use it to directly assess potential departures from the basic model for our discussion of earnings inequality.
**Figure A.18 – Comparison to Piketty and Saez (IRS data)**

(A) 90th Percentile

(B) 95th Percentile

(C) 99.5th Percentile

(D) 99.9th Percentile

Notes: Piketty and Saez (2003) data is based on Table B3 in http://eml.berkeley.edu/~saez/TabFig2014prel.xls. All values are adjusted for inflation using the PCE price index. For SSA data, only individuals in firms with at least 20 employees are included. Only full-time individuals aged 20 to 60 are included in all SSA statistics, where full-time is defined as earning the equivalent of minimum wage for 40 hours per week in 13 weeks. Individuals and firms in public administration or educational services are not included in SSA data.

A few additional technical aspects are worth highlighting. The linear age component is not separately identified when worker effects and year effects are present. If one simply drops the linear age effects, the estimated variance of the worker effects is biased. Instead, we follow CHK and normalize age by subtracting and dividing by 40. Since at age 40 the marginal effect of age on earnings is approximately equal to zero, the estimated worker effects and their variance are unbiased.\(^{29}\) However, as is well known, there is still a finite sample bias in estimates of

\(^{29}\)The age-earnings gradient in SSA data flattens out around age 40. The worker effect is biased.
**Figure A.19 – Regression residuals by firm fixed effect decile**

(A) 1980-1986

(B) 2007-2013

(C) Change from 1980-1986 to 2007-2013

Notes: Calculations based on SSA data. See section 4.2 for sample selection criteria. Within each firm FE decile group, worker FE deciles are order from left to right from 1 to 10.

because it absorbs the time-invariant effect of age (i.e., age at start of the sample, which is effectively
Figure A.20 – Distribution of workers among firm FE deciles, by firm size

(A) 1980-1986

(B) 2007-2013

(C) Change from 1980-1986 to 2007-2013

Notes: Calculations based on SSA data. See section 4.2 for sample selection criteria. Within each firm size group, firm FE deciles are order from left to right from 1 to 10.

— a cohort effect. Note that for the analysis of changes in the variance of worker effects over time, the
Figure A.21 – Distribution of workers among worker FE deciles, by firm size

(A) 1980-1986

(B) 2007-2013

(c) Change from 1980-1986 to 2007-2013

Notes: Calculations based on SSA data. See section 4.2 for sample selection criteria. Within each firm size group, worker FE deciles are order from left to right from 1 to 10.

normalization has no effect on the trend as long as the age distribution of the population and the return
**Figure A.22** – Conditional Distribution of Workers among Deciles of Worker and Firm Fixed Effects

(A) 1980-1986

(B) 2007-2013

(C) Change from 1980-1986 to 2007-2013

Notes: Calculations based on SSA data. See section 4.2 for sample selection criteria. Firm and worker fixed effects from our AKM estimation sorted into deciles. Statistics are computed with respect to the total number of observations in each firm FE decile, as opposed to the full population. Firm fixed effect deciles are computed with respect to the distribution of firms. Within each firm FE decile group, worker FE deciles are ordered from left to right from 1 to 10.
var_j(ψ^j) and var_i(θ^i) because of sampling error in the estimated worker and firm effects.

In addition, the estimate of the covariance term (cov(θ^i, ψ^j)) is likely to be downward biased, because the sampling error in the worker and firm effects are negatively correlated. We do not attempt to construct bias-corrected estimates of these components. Instead, we follow the literature and focus on trends in the estimated moments assuming that the bias from sampling errors is similar over time. Finally, firm effects are identified up to the difference with respect to an omitted reference firm. Hence, one can only obtain comparable estimates of firm effects for firms that are connected by worker flows. Following AKM and CHK, we estimate equation (3) on the greatest connected set of workers, which in our case comprises close to 98% of all observations (see Table C1).

C.2 Model Fit

Table C1 shows basic characteristics for the full sample as well as for observations in the connected set, separately for each of our five time periods. In the following, we will focus our discussion on men. Unless otherwise noted, the results for women are similar. For space reasons, the results for women are in an appendix. Table C1 shows that in all five periods, approximately 98% of workers are in in the greatest connected set. As a result, the mean, median, and standard deviation of earnings in the connected set are very similar to the overall sample. If one compares the number of observations with the number of workers, one obtains that the average worker is in the sample about 5 of 7 years in each period. This number is very similar to numbers reported by CHK (Table I) for full-time men in Germany.

Table C2 displays basic statistics from the estimation. The table delivers a snapshot of the basic findings, as well as important diagnostic checks. In terms of basic findings, the table shows how the standard deviation of worker effects has risen over time, especially in the early 1980s. The standard deviation of firm effects has remained stable. In contrast, the correlation of worker and firm effects rose almost five fold from our first period, 1980-1986, to our last period, 2007-2013. The table also shows that the RMSE has remained stable, and has at best declined somewhat over time. If the rise in sorting of workers to firms had resulted from an increasing role of complementarities (i.e., match effects), we would have expected the goodness of fit of the model without match effects to decline over time. Instead, the RMSE drops at the same time as the variance of earnings increases. As a result, the adjusted R^2 increases from 74.1% in 1980-1986 to 81.2% in 2007-2013.

While the goodness of fit based on worker and firm effects and age is quite high, at around 80%, there is room left for additional components. To check whether adding a match-specific component would substantially increase the fit of the model, the bottom of the table shows basic statistics of a model that also allows for a match effect (m_{ij}). Not surprisingly, allowing

to age are roughly stable over time. The firm effects is not affected by the normalization. The covariance of worker and firm effects may be affected insofar as workers are sorted into firms by age.

Andrews et al. (2008) show that the degree of negative correlation declines with the number of movers that is used to identify firm effects. We indeed find that the level of the covariance rises with our sample size. However, the gradient over time is unaffected.

The correlation of observable worker characteristics (mainly age) with worker and firm effects has a U-shaped pattern—declining to a low point during the economic book of the late 1990s, and returning to similar levels by the end of the period.
for a match effect reduces the RMSE and increases the adjusted $R^2$, by about the same amount each period, to 82 – 87%. However, the standard deviation of match effects declines somewhat over time. As noted by CHK, this is consistent with an interpretation of the match effects as uncorrelated random effects. If instead they were specification errors caused by incorrectly imposing additivity of the person and establishment effects, one would expect the standard deviation of match effects to rise and the relative fit of the AKM model to deteriorate over time as the covariance of worker and firm effects increases in magnitude.

As additional check on the appropriateness of the basic AKM specification of model (3), we examined average regression residuals for different groups of worker and firm effects. Violations of the separability assumptions in the AKM model would likely cause large mean residuals for certain matches, say, where highly skilled workers are matched to low-wage establishments. To search for such potential interactions, we followed CHK and divided the estimated person and establishment effects in each interval into deciles, and computed the mean residual in each of the 100 person firm decile cells.

Figure A.19b shows the mean residuals from the cells using data from period 2007–2013. For most cells, the mean residual is very small, below 0.02, and shows few systematic patterns. Only for cells with either low worker effects or low firm effects do residuals appear larger. It is interesting to note that this pattern is quite similar to those found by CHK (Figure VI), who report larger mean residuals for the lowest worker and firm effect groups. Hence, in both Germany and the U.S. separability appears a good description for all worker and firm groups but for the bottom end.\footnote{Not surprisingly given the presence of labor supply effects, the mean residuals in Figure A.19 are on average larger than those shown CHK (Figure VI).} Figure A.19c shows the change in mean residuals within cells over time. The changes are of opposite signs of the deviations in A.19b, implying that the absolute magnitude of deviations has declined over time. Hence, overall, the goodness of fit of the model has improved from the first to the last period in our sample.

Our last diagnostic assesses the ability of the model to explain earnings changes at job changes. If the model is correctly specified, on average workers changing from one firm to another should experience earnings changes corresponding to the estimated firm effects. To implement this comparison, we used our data to perform event-study analyses of the effect of job mobility on earnings akin to those shown in CHK (Figure VII). As in CHK, we divided firms into quartiles according to both their average wage and their firm effects, and recorded the mean earnings of workers moving between the four firm-type classes in the years before and after the job change. One complication is that we do not observe when in a given year a worker leaves his initial employer, and whether he joins his new employer in the same year or at some point in the adjacent year. To deal with the fact that we do not know the specific time of the move, we followed workers from two years before the year $t$ in which we observe the move (i.e., from year $t - 2$ to $t - 1$), to two years after the year succeeding the move (from year $t + 2$ to $t + 3$). To further try to approximate transition between “full-time” jobs, we only look at workers who remained at the firm in the two years before and two years after the move. Since we are following workers for six years, we adjust earnings for flexible time trends. The results are shown in Figures A.23 and A.23, for firm classes based on firm fixed effects and firm average wages, respectively. The Results are discussed in the main text in Section 4. Overall, we conclude that despite the fact we are modeling information on annual earnings
rather than daily or hourly wages, our model delivers a good approximation of the underlying earnings process.

**Figure A.23** – Event study of change in mean earnings for job changers

(A) Firms ranked by earnings: 1980-1986

(B) Firms ranked by earnings: 2007-2013

Notes: Calculations based on SSA data. See section 4.2 for sample selection criteria. For an explanation of the methodology see Section 4.2 and Appendix C.2. For all observations main job is the same in years 1, 2, and 3 and then switches to a new main job for years 4, 5, and 6. The shaded region marks the possible years of the job switch. Mean earning quartiles are weighted by worker-years and calculated in years 2 and 5. Mean earnings are computed as “leave-out” means, i.e. for each individual, mean firm earnings are computed over all employees except the reference employee. Log earnings are detrended by subtracting the time-varying observable AKM component from each observation.

**C.3 Patterns of Sorting By Firm Earnings, Firm Size, and Industry**

Tables 2 and 3 have shown that the substantial between-firm component of the rise in earnings inequality in the United States from the early 1980s to today can be attributed almost entirely to sorting (a rise in the correlation of worker and firm effects) and segregation (a rise in the variance of mean worker effects between firms). We have also found that these patterns are particularly pronounced for moderately sized firms (i.e., for employment size less or equal to 1000). In this section, we will use our estimated worker and firm effects from implementing
equation (3) to assess how workers are sorted into high-wage firms and large firms, and how this has changed over time. We will also describe the changing patterns of firm and worker effects by firm size and industry.

To learn more about the pattern of sorting, the first two panels of Figure 7 displays the joint distribution among deciles of worker and firm effects in 1980-1986 and 2007-2013. The cross-sectional sorting patterns displayed in the figure are striking. Consider first the early 1980s shown in Figure 7a. One can see that most workers are in medium to high fixed-effect firms. Yet, lower fixed-effect workers are over-represented at lower fixed-effect firms; workers with fixed effects in the middle range are over-represented at middle to high fixed-effect firms; and high fixed-effect workers are over-represented at high fixed-effect firms. However, one also sees that low to medium fixed-effects firms have modes at both low and high fixed effects workers, presumably reflecting a distribution of lower-skilled production workers and managerial employees.

The distribution for the years 2007–2013 displayed in Figure 7b show these patterns have changed substantially over time. Figure 7c shows the net change of density of the two distributions at corresponding deciles. Overall, there has been a substantial shift in the distribution away from the two highest firm categories towards middle to lower fixed-effect firms. Yet, this shift did not occur uniformly across worker groups. It is the middle of the worker fixed-effect distribution that predominantly left high-wage firms, such that high-wage workers are now over-represented at the top firms. This pattern is augmented by a move of the highest fixed-effect workers to higher-paying firms.

To better display the relative patterns of change within firm categories, Figure A.22 shows the conditional distribution of workers within firm effect deciles for our first and last time period and the change over time. Figure A.22c displays the change in the pattern of sorting most clearly—middle-wage workers move to the middle-and high-wage workers move to the top. The only exception to this pattern is the lowest decile of firm effects. Yet, Figure 7 shows this group contains few workers to begin with (Figure 7a), and exhibits very little net change (Figure 7c).

Figures 7 and A.22 confirm the evidence from the variance decomposition that sorting has increased, and show which workers and firms appear most affected. A striking finding is that the incidence and composition of workers at high-wage firms has been changing substantially. Since high-wage firms are likely to be in part large firms, and we have found large firms to play a special role in the evolution of inequality, we use our data to examine the incidence of worker and fixed effects separately by firm size. These results are shown in Figures A.20 and A.21 for three firm size groups (firms with number of workers in range 1 – 100, 101 – 9999, and 10,000+).

From Figure A.20 it is clear that on average, high-wage firms tend to be larger. However, over time, Figure A.20c shows that large employers have experienced a substantial shift out of high-wage firms to middle and lower-wage firms. Figure A.21 shows that among larger firms, the decline was accompanied by an increase in the incidence of high wage workers at larger firms. In addition, especially employers with more than 10,000 employees saw a reduction in

---

33Note that the definition of the deciles differ between the two time periods. Yet, since the distribution of firm effects has changed little, the deciles of firm effects are roughly comparable over time.
workers in the middle of the worker fixed-effects distribution. Hence, this confirms that larger firms have become, on average, workplaces that pay less and employ a more unequal set of workers.

To examine potential differences in sorting patterns, we have also examined the joint distribution of firm and worker fixed effects within each size class. These figures are displayed in our appendix. The results show that the pattern of sorting is quite similar among our two larger firm size classes, and reflects the pattern shown in Figure 7 – there is a substantial net shift in the mass of workers from high-wage firms to middle-wage firms. The bulk of this shift is comprised of middle-wage workers. In contrast, high-wage workers have left middle-wage firms to move to the top firms. In contrast, the distribution of low-wage workers has changed less. These results corroborate our finding from Table 2 that the differences in the sources of inequality growth by firm size is not the between-firm component, whose levels evolve similarly, but rather the within-firm component of inequality.

We have also examined the pattern of marginal distributions of firm and worker effects by one-digit industry. These figures are again contained in our appendix. The results show that the large decline in the incidence of employment in higher wage deciles tends to be concentrated in manufacturing. Employment at high-wage manufacturing firms is increasingly replaced by employment in middle-wage service firms. In terms of workers, middle-wage workers have again shifted out of manufacturing, and moved to services. Yet, services has also received an increasing proportion of high-wage workers, with low-wage workers increasingly moving to firms with unknown industry affiliation. These are likely to be disproportionately new employers, which might be likely to have low firm fixed effects.

Overall, the findings from the figures corroborate and strengthen our core results from the detailed variance composition in Table 3. There is a clear pattern of increasing sorting of higher-wage workers into higher-wage firms over time. In particular, from the early 1980s to today, high-wage firms appear to lose middle-wage workers to middle-wage firms, and in turn gain more high-wage workers. These patterns partly correspond to shifts between firm-size classes. Fewer middle-wage workers work at very large employers, at the same time as these employers are increasingly composed of lower-wage firms. Yet, within firm-size classes the patterns of sorting is similar as for the full sample, and characterized by a substantial shift between firm-size classes and substantial redistribution of workers. Overall, these findings hint at a substantial reorganization of U.S. businesses over the last 40 years. This reorganization has had profound consequences for both the level and the nature of earnings inequality.