FIRMING UP INEQUALITY

Online Appendix

Jae Song

David J. Price

Fatih Guvenen

Nicholas Bloom

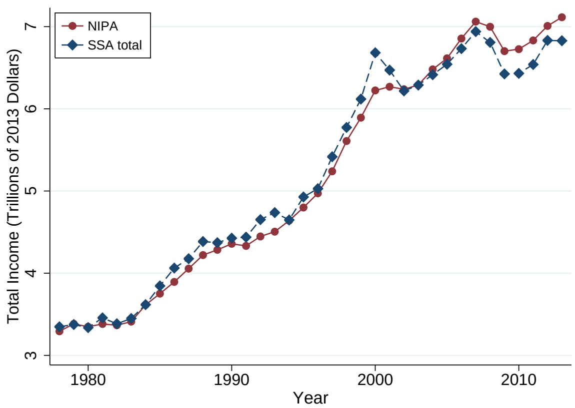
Till von Wachter

Contents

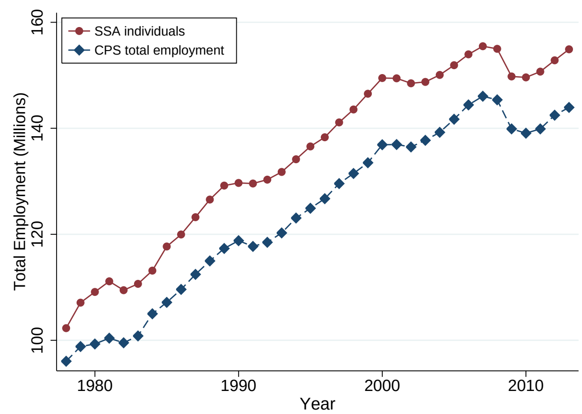
A Additional Figures

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

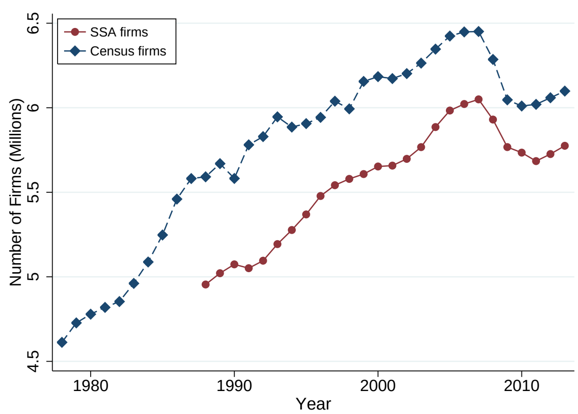
(a) Total earnings



(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.

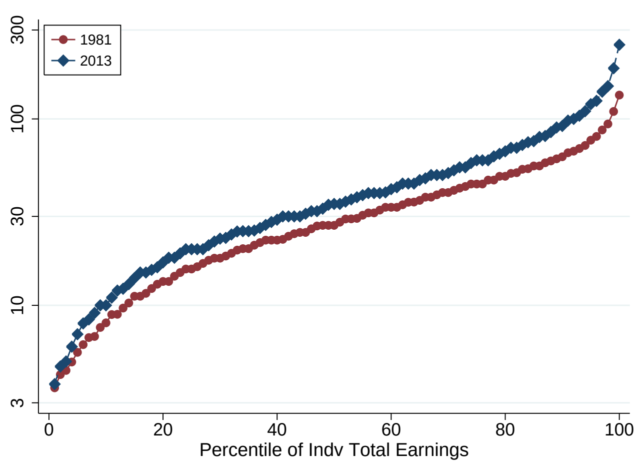
Figure A.2 – Comparing earnings variance in SSA and CPS data

(A) Levels

(B) Change since 1981

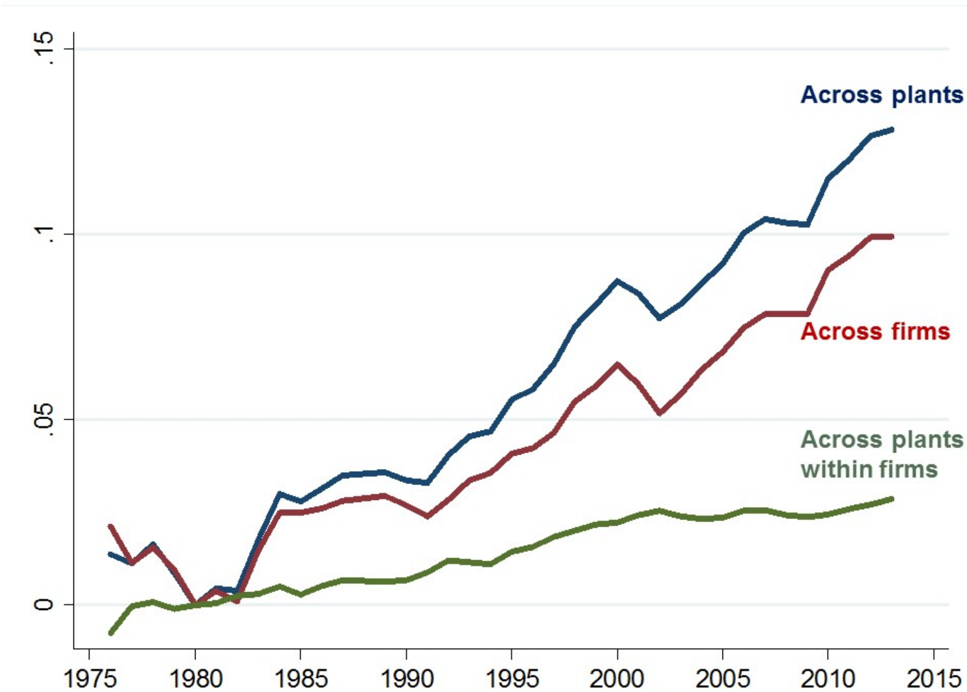
Notes: Only employed individuals aged 20 to 60 are included in all statistics, where "employed" is defined as earning the equivalent of minimum wage for 40 hours per week in 13 weeks. Individuals and firms in public administration or educational services are not included. Only firms and individuals in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

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



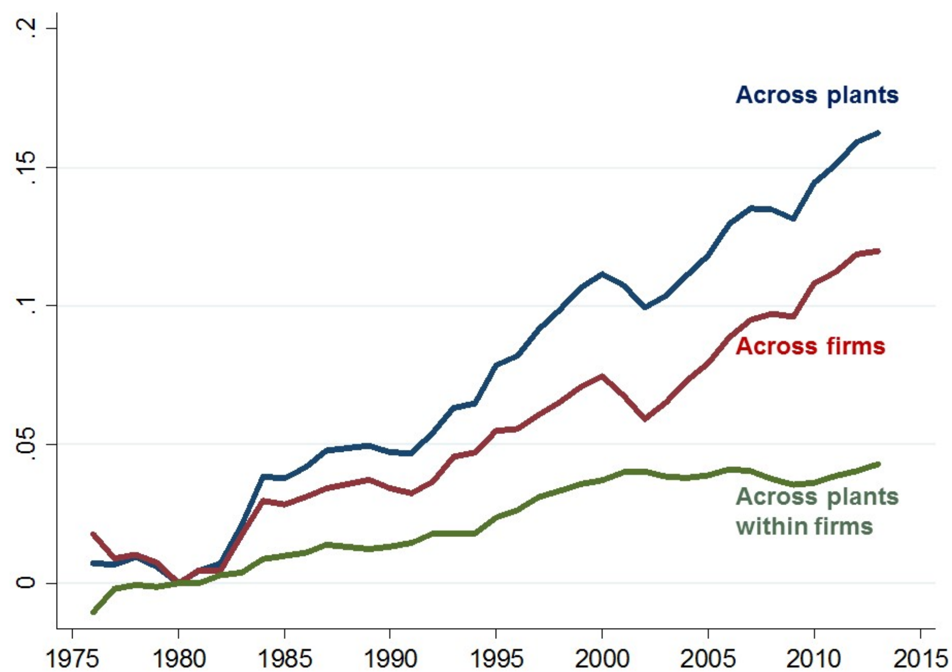
Notes: For each percentile, statistics are based on the minimum earnings among individuals in that percentile of earnings in each year. All values are adjusted for inflation to 2013 dollars using the PCE price index. Only employed individuals aged 20 to 60 are included in all statistics, where "employed" 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 – Variance of log(wages) across establishments and firms - firms with 20+ workers



Notes: Plots the employment weighted variance of log wages normalized to 0 in 1980. Source Census Longitudinal Business Database 1976-2013. All establishments and firms with positive employment and wages. All sectors excluding education (SIC codes 8200 to 8299) and public administration (sic codes 9000 to 9899). Establishments are dropped if their average wage (defined as total wages/employment) is above $250,000 or below $12,180 (minimum wage for 35 hours a week for 48 weeks a year) in 2013 dollars. Establishments only from firms with 20+ employees.

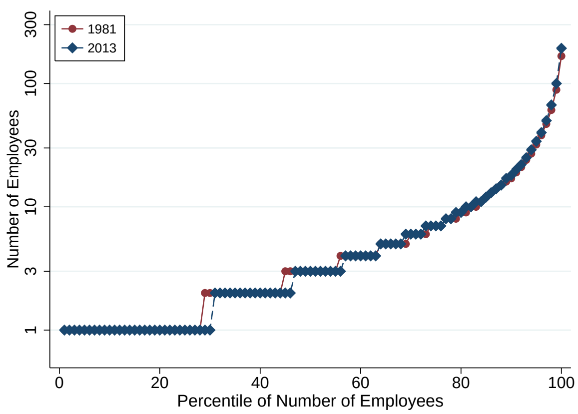
Figure A.5 – Variance of log (wages) across establishments and firms - establishments in mega firms



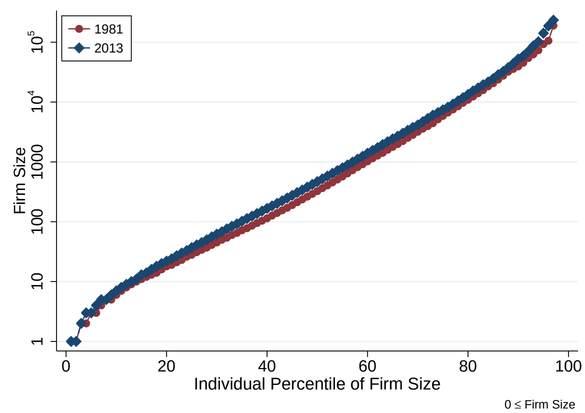
Notes: Plots the employment weighted variance of log wages normalized to 0 in 1980. Source Census Longitudinal Business Database 1976-2013. All establishments and firms with positive employment and wages. All sectors excluding education (SIC codes 8200 to 8299) and public administration (sic codes 9000 to 9899). Establishments are dropped if their average wage (defined as total wages/employment) is above $250,000 or below $12,180 (minimum wage for 35 hours a week for 48 weeks a year) in 2013 dollars. Establishments only from firms with 10,000+ employees.

Figure A.6 – Cumulative firm size distribution

(a) Unweighted



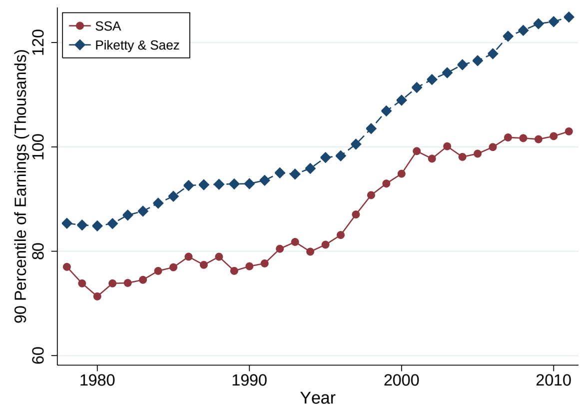
(b) Weighted by employment



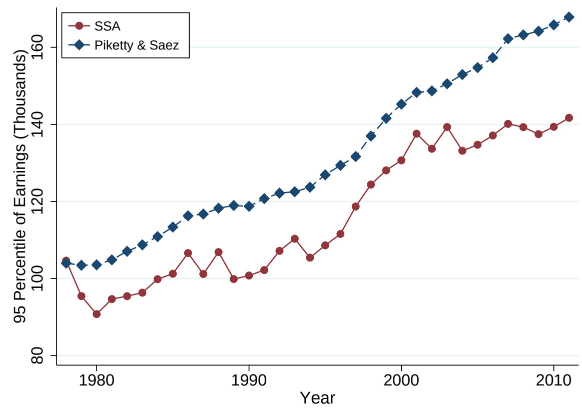
Notes: Only employed individuals aged 20 to 60 are included in all statistics, where "employed" is defined as earning the equivalent of minimum wage for 40 hours per week in 13 weeks. Individuals and firms in public administration or educational services are not included. Both graphs are inverse cumulative distribution functions. Figure A.6a shows the fraction of firms below a given size; Figure A.6b shows the fraction of individuals at firms below a certain size. For disclosure reasons, Figure A.6b does not report the top 3 percentiles.

Figure A.7 – Comparison to Piketty and Saez (IRS data)

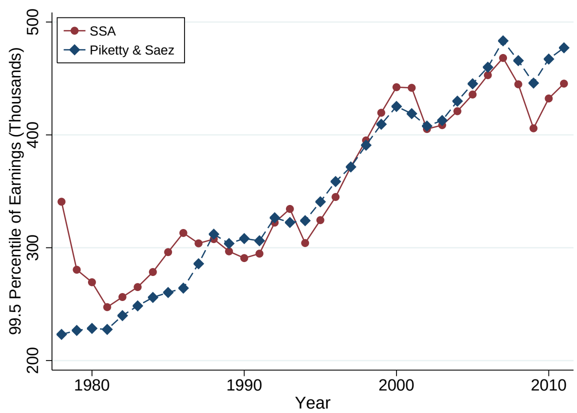
(a) 90th Percentile



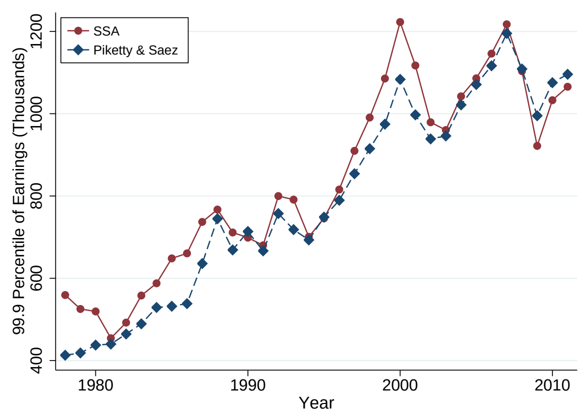
(b) 95th Percentile



(c) 99.5th Percentile



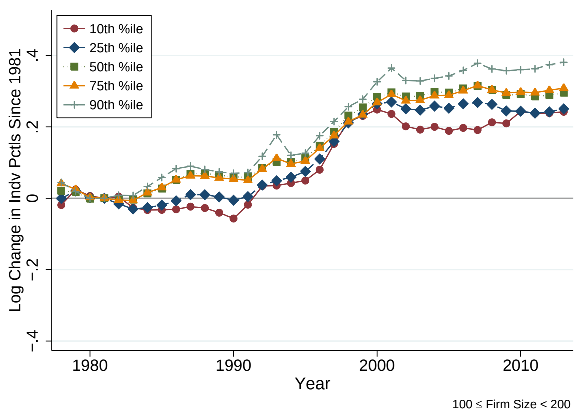
(d) 99.9th Percentile



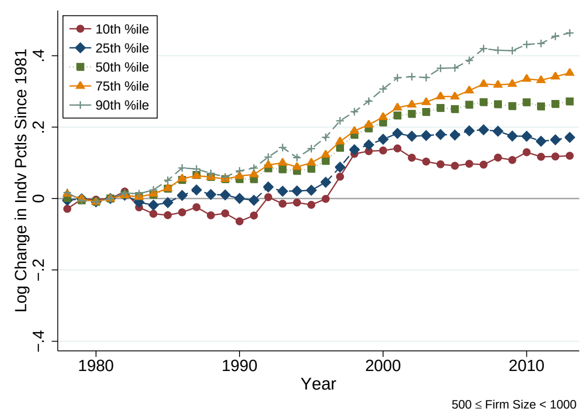
Notes: Piketty and Saez (2003) data is based on Table B3 in http://eml.berkeley.edu/~saez/TabFig2014prel.xls. All values are adjusted for inflation to 2013 dollars using the PCE price index. For SSA data, only individuals in firms with at least 20 employees are included. Only employed individuals aged 20 to 60 are included in all SSA statistics, where "employed" 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.

Figure A.8 – Alternative Version of Figure VI: Change in Percentiles of Annual Earnings Distribution for Workers Grouped by Employer Size. Workers are first assigned to samples defined by their employers' size, and percentiles are computed for each sample separately. Select samples are shown to save space.

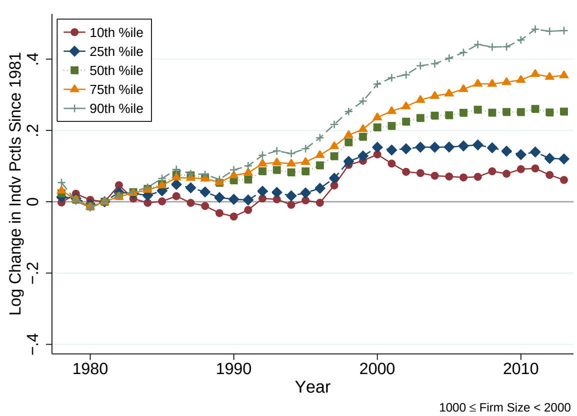
(A) Workers at Firms with 100 to 200 employees



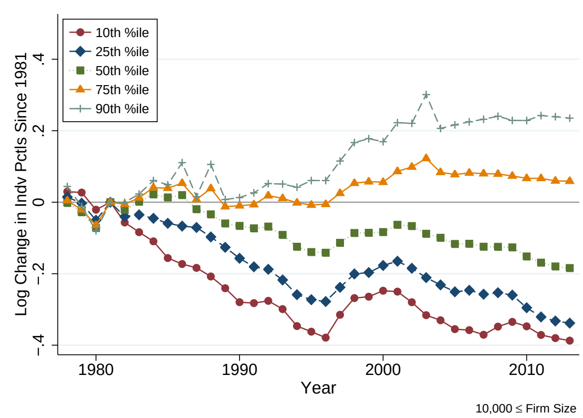
(B) Workers at Firms with 500 to 1000 employees



(C) Workers at Firms with 1,000 to 2,000 employees



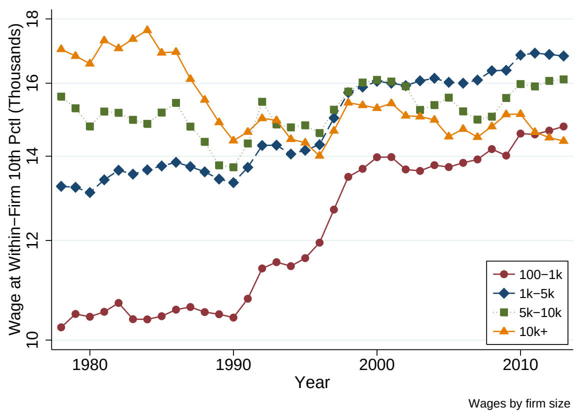
(D) Workers at Mega Firms



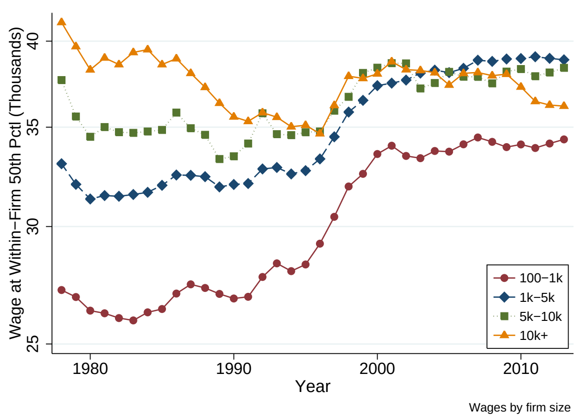
Notes: Only firms and individuals in firms with the given number of employees are included. Only employed individuals aged 20 to 60 are included in all statistics, where "employed" is defined as earning the equivalent of minimum wage for 40 hours per week in 13 weeks. Individuals and firms in public administration or educational services are not included.

FIGURE A.9 – Figure VI, Non-scaled version. Levels of Percentiles in Thousands of Dollars. All other details same as Figure VI.

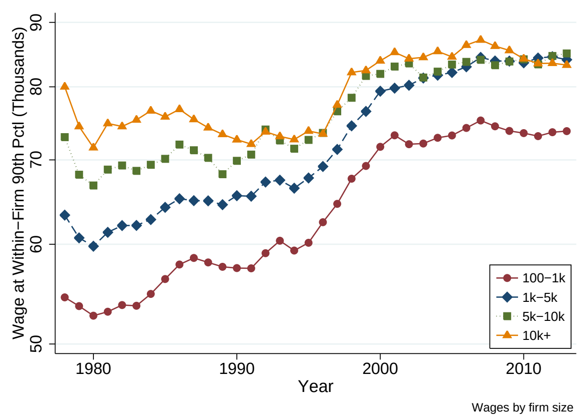
(A) 10th Percentile



(B) 50th Percentile



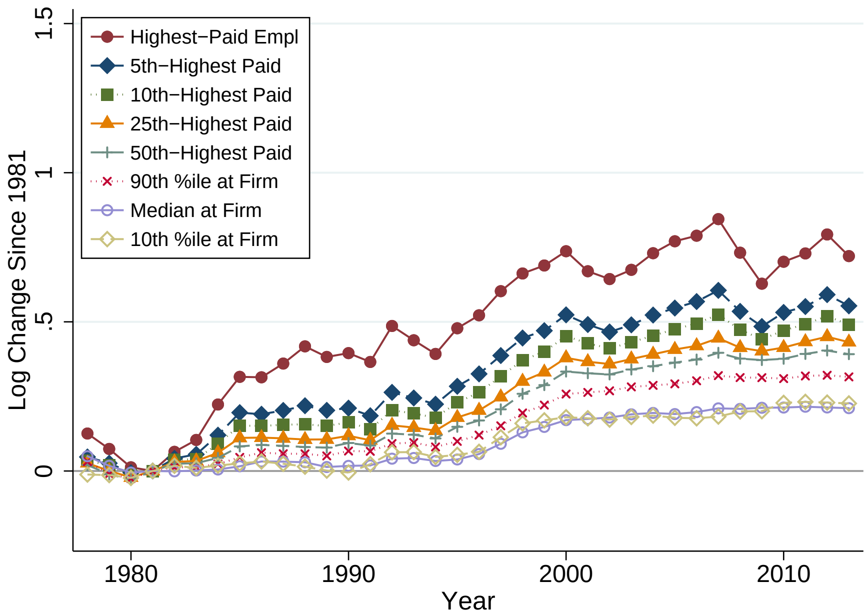
(C) 90th Percentile



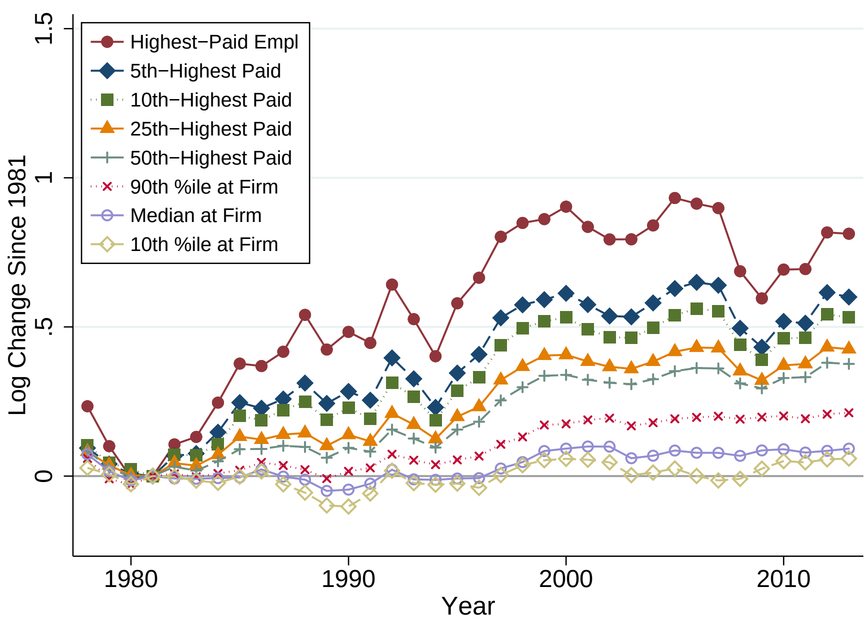
Notes: Only firms and individuals in firms with the given number of employees are included. Only employed individuals aged 20 to 60 are included in all statistics, where "employed" is defined as earning the equivalent of minimum wage for 40 hours per week in 13 weeks. Individuals and firms in public administration or educational services are not included.

Figure A.10 – Change in within-Firm Distribution of Annual Earnings: Other Sizes

(a) Workers at Firms with 1,000 to 5,000 employees



(b) Workers at Firms with 5,000 to 10,000 employees



Notes: Only firms and individuals in firms of the listed size are included. Only employed individuals aged 20 to 60 are included in all statistics, where "employed" 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 to 2013 dollars using the PCE price index.

B Data Procedures

B.1 Social Security Administration Data

As noted in Section II, 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 concretely, we consider an individual to be employed 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.)[1] 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). 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 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

[1] Note that the worker-firm fixed effect model instead imposes the restriction of 520 hours at the 2013 minimum wage, adjusted for inflation with the PCE.

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. 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 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 VI and VIII, 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. Finally, we adjust all dollar values in the data set to be equivalent to 2013 dollars with the Personal Consumption Expenditure (PCE) price index.[2]

There are several differences between the data used in the worker-firm fixed effect model (CHK model) and the rest of the text. The most important difference is that the CHK model only includes data on men. Men are included if they earned the equivalent of at least 520 hours at the 2013 minimum wage, adjusted for inflation with the PCE; in the rest of the text, the minimum earnings threshold was 520 times the contemporaneous minimum wage. To estimate the model, we included men in all firms, regardless of the firm's size or industry. We then report results based on the same sample as the rest of the paper: firms, and people in firms, with at least 20 total (male and female) employees, who are not in public administration. (The only exceptions are Tables A.5 and A.6, which report summary statistics based on all observations used to calculate fixed effects.) Thus if someone moved from a 5-person firm to a 50-person firm, that move would be part of the sample used to calculate fixed effects, but only data from the second year would only be included in our estimation results. If they stayed at the same firm and the firm grew from 5 to 50 people, both years would be included in the sample used to calculate fixed effects, but only data from the second year would be included in estimation results.

[2]http://research.stlouisfed.org/fred2/series/PCEPI/downloaddata?cid=21

B.2 Current Population Survey Data

We use 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 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.

C Further Analysis on Earnings Inequality within and between Firms

C.1 Coworkers of Individuals across the Entire Distribution

Figure A.11 provides information similar to Figure III but follows Juhn et al. (1993) and many related papers in showing the change between 1981 and 2013 for each percentile in the earnings distribution. It is important to realize that this graph, unlike Figure IV, is not a counterfactual analysis; instead, it shows how the relationship between individual earnings and coworker earnings changed at different points in the earnings distribution. Understanding the average earnings of coworkers is important for understanding how workers might perceive inequality, among other reasons, but it cannot tell us, for example, how inequality would have been different if between-firm differences in average earnings had been unchanged.

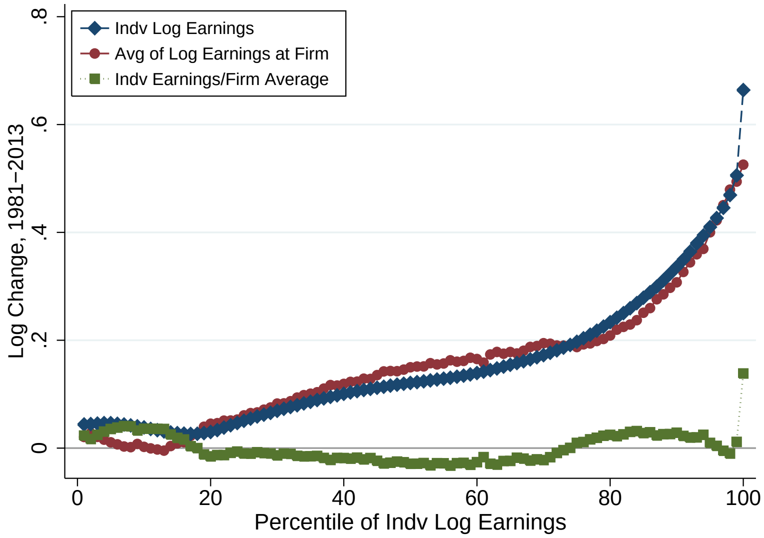
We start with the dashed line marked with diamonds (labeled "Indv Total Earnings"), which shows the increase in log earnings between 1981 and 2013 within each percentile of the earnings distribution.[3] So, for example, we see that between 1981 and 2013, the 50th percentile of earnings has increased 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.[4]

To assess how average earnings per worker of employers of workers in each percentile of the earnings distribution has changed, we repeat an exercise similar to that for Figure IIIb.

[3]This graph is closely related to the difference between the 2013 and 1981 lines in Figure Ia, which shows percentiles of earnings in each year. The only difference results from the fact that Figure Ia shows the minimum earnings within each percentile, while Figure IV is based on average log earnings in each percentile.

[4]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 IV, we 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 experienced a substantial increase.

Figure A.11 – Change in Inequality of Annual Earnings across Percentiles from 1981 to 2013



Notes: Only firms and individuals in firms with at least 20 employees are included. Only employed individuals aged 20 to 60 are included in all statistics, where "employed" 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 to 2013 dollars using the PCE price index.

For a given percentile, we take firm average earnings and average it across all the employers of workers in that percentile separately in both 1981 and 2013, and then take the difference between the years (shown in Figure A.11 as a line marked with circles, labeled "Avg of Log Earnings at Firm"). The upward slope of this "Avg of Log Earnings at Firm" line indicates that the firms of high-earnings individuals now have higher average earnings than firms of high-earnings individuals in 1981, while firms of low-earnings individuals had roughly the same average earnings as firms of low-earnings individuals in 1981.

Finally, the "Indv Earnings/Firm Average" line (marked with squares) reports changes in the ratio of own log earnings to firm average log earnings for those at different points in the individual distribution.[5] Particular care should be given to the interpretation of this line, which is almost flat across all percentiles. Taken together, this graph indicates that although highly paid individuals are now being paid much more than highly paid individuals were in 1981 (as evidenced by the "Indv Total Earnings" line), they are also at firms where their coworkers are being paid better (the "Avg of Log Earnings at Firm" line). Thus their earnings relative to that of their coworkers has barely changed since 1981. (For poorly paid individuals, own

[5]Note that this "Individual/Firm" line will be mechanically equal to the difference between the "Individual" line and the "Firm" line. Also, the "Individual/Firm" line's average taken over all percentiles must be zero.

earnings and their firm's average earnings changed little in the past few decades, so the ratio is also mostly unchanged.)

The difference between the results of Figures IV and A.11 points to another core result of the paper. As discussed in detail in Section IV, the fact that the average earnings of coworkers throughout the distribution has increased proportionally to the rise in individual earnings is partly explained by the fact that higher-wage workers are increasingly working at higher-wage firms and are increasingly working with other higher-wage workers.

C.2 Inequality at the Top of the Earnings Distribution

C.2.i The Top 1% of Earners, Relative to their Firms

Much of the recent policy and media attention around inequality has focused on the rising share of earnings going to the top 1%. One interesting question in this context is to what extent these very top earners have pulled away from their coworkers in the same firm as opposed to experiencing rising earnings together with the rest of their firm (i.e., the between versus within question). To shed light on this question, in Figure A.12 we plot the analog of Figure A.11, but this time focusing entirely on the top 1% and splitting it into 100 quantiles of 0.01% each. (With about 70 million people in the full sample per year, each 0.01% represents about 7,000 people on average.)

We see in Figure A.12 that up until about the 99.5% point—which is an earnings threshold of around $450,000 in 2013 (see Figure Ib)—increases in individual earnings from 1981 to 2013 within each percentile point have been matched almost fully by the increases in earnings of their coworkers. 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 earnings by 73 log points (107%), generating a 30 log point gap.[6] Thus, according to this metric, earners in the top 0.5% have seen substantial earnings 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, and engineers, among others (Guvenen et al. (2014)).

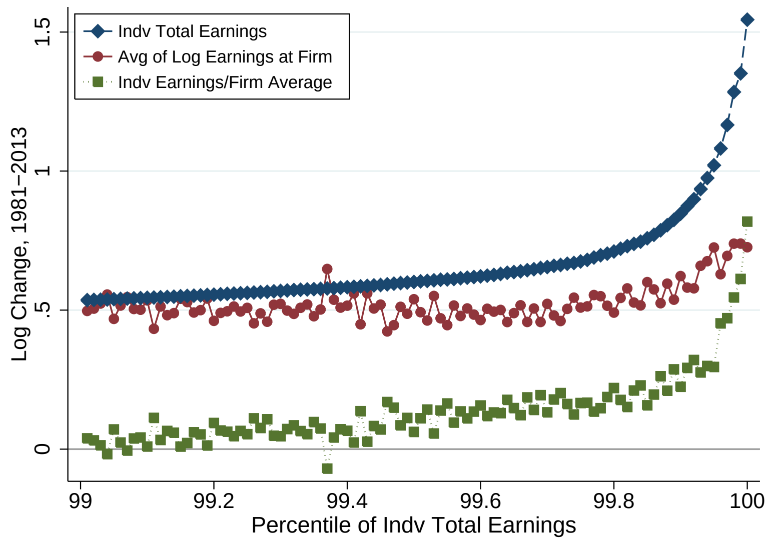
C.2.ii Top Earnings Share

In fact, our data allow us to speak about the share of earnings going to individuals at the top of their firms. Piketty and Saez (2003) describe the increasing fraction of income that is going to the top few percentiles of the income distribution. In this section, we note that, while those at the top of their firms are earning a greater fraction of earnings in the overall economy, their share within the top 1% and top 0.1% of economy-wide earnings has changed little in the past three decades.

People at the top of their firms in 2013 generally receive a greater share of economy-wide earnings than those at the top in 1981. In firms with at least 20 employees, the top 1% within

[6] 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.

Figure A.12 – Rise in Inequality of Annual Earnings between 1981 and 2013 among Top 1% of Earners



Notes: Only firms and individuals in firms with at least 20 employees are included. Only employed individuals aged 20 to 60 are included in all statistics, where "employed" 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 to 2013 dollars using the PCE price index.

Table A.1 – Percentage of Top 1% Earnings in the Economy Going to Those at the Top of Their Firms

	1981	2013
Top-paid person at firm	23%	17%
Among top five at firm	42%	37%
Among top 1% at firm	55%	50%

Notes: Statistics are reported for all people who are in firms with at least 20 employees. Only employed individuals aged 20 to 60 are included in all statistics, where "employed" 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 show, of all earnings going to the top 1% overall, how much went to the top-paid person at firms; those who are among the top five top-paid employees at their firm; and those who are among the top 1% at their firm.

each firm took home 8.4% of total earnings, an increase from 4.7% in 1981. (In mega firms, the top 1% took home 7.1% in 2013 and 4.7% in 1981.) However, these gains generally match those of other high earners, as shown in Table A.1. For example, of all earnings that went to those in the top 1% overall, those who were the top-paid person at their firm (most likely the

Table A.2 – Percentage Who Are Top-Paid Person at Firm

	1981	2013
All individuals	0.71%	0.62%
Top 1%	18%	11%
Top 0.1%	36%	20%

Notes: The percentage of people who are the top-paid person in their firm, among all individuals in firms with at least 20 employees. Only employed individuals aged 20 to 60 are included in all statistics, where "employed" 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. Rows show the fraction who are top paid among all individuals; those who are in the top 1% of top earners; and those in the top 0.1% overall.

CEOs) earned 23% in 1981 but only 17% in 2013.

Indeed, high earners are now less likely to be the top-paid person at their firms, as shown in Table A.2. The fraction of the top 1% who were the top-paid person at their firm declined from 18% in 1981 to 11% in 2013, while the fraction of those in the top 0.1% who were the top-paid person declined from 36% to 20%. This decline is only partially due to the increasing size of firms: for comparison, the total fraction of people who are the top-paid person at their firm decreased from 0.71% to 0.62%. Rather, it is because highly paid employees are more likely to work with other highly paid employees.

These results mirror those found by Bakija et al. (2012). Using IRS tax data, they found that, although nonfinancial executives, managers, and supervisors in the top 1% and 0.1% earned an increasing fraction of overall income, their total income relative to others in the top 1% was mostly flat.

Table A.3 – Complete Robustness Checks on Variance Decomposition - Part 1

	Total Var, 1981	Between-Firm Var, 1981	Total Var, 2013	Between-Firm Var, 2013	Total Var Increase	Frac Increase Between
Baseline sample	0.652	0.222	0.846	0.357	0.194	0.694
Any number of empl	0.691	0.272	0.852	0.387	0.161	0.71
20-10k workers	0.651	0.206	0.835	0.36	0.184	0.837
10k+ workers	0.552	0.164	0.873	0.348	0.32	0.577
Avg 5-year earnings	0.69	0.207	0.861	0.314	0.171	0.629
Min earn = 260 x min ‘wage	0.823	0.284	1.03	0.434	0.206	0.73
Min earn = 1040 x min wage	0.479	0.159	0.658	0.272	0.179	0.628
Min earn = 2080 x min wage	0.33	0.101	0.48	0.179	0.151	0.518
Min earn based on 2013 min wage	0.641	0.218	0.846	0.357	0.205	0.676
Excluding top 1%	0.607	0.21	0.771	0.323	0.164	0.694
Excluding top 1% in each firm	0.619	0.227	0.804	0.359	0.185	0.717
Excluding top 5%	0.537	0.183	0.664	0.265	0.127	0.644
Excluding top 5% in each firm	0.577	0.23	0.744	0.36	0.167	0.776
Excluding top-paid person in each firm	0.64	0.228	0.835	0.36	0.195	0.681
Excluding 5 top-paid people in each firm	0.627	0.243	0.815	0.369	0.188	0.67
Women only	0.485	0.164	0.74	0.306	0.255	0.555
Men only	0.619	0.195	0.89	0.389	0.271	0.718
Ag/Mining/Construct/Oth	0.633	0.169	0.822	0.34	0.19	0.905
Manufacturing	0.576	0.191	0.676	0.251	0.1	0.605
Utilities	0.444	0.111	0.611	0.207	0.167	0.577
Trade	0.678	0.173	0.797	0.275	0.119	0.858
FIRE	0.589	0.12	0.848	0.274	0.259	0.593
Services	0.676	0.231	0.832	0.333	0.157	0.652

Notes: Only firms and individuals in firms with at least 20 employees are included. Only employed individuals aged 20 to 60 are included in all statistics, where “employed” 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. “Total Var” indicates total variance of earnings in a given year, and “Between-Firm Var” indicates total between-firm variance in that year. “Total Var Increase” denotes the increase in variance between 1981 and 2013, while “Frac Increase Between” denotes the fraction of that increase in variance accounted for by an increase in between-firm variance. Statistics in rows labeled “Demean by” include earnings that are demeaned within a given group before all variances are calculated. Statistics in rows with showing numbers of employees are limited to individuals in firms with that number of employees. Demeaning by “firm size category” includes each category listed starting with “Size:” in this table. “Avg 5-year earnings” performs the same analysis but uses all earnings over 5 years, 1981-1985 to 2009-2013, to calculate statistics. “Min earn =” uses different minimum earnings thresholds. “Min earn based on 2013 min wage” uses the 2013 minimum wage adjusted for inflation, rather than the minimum wage in the given year. “Excluding top...” exclude top people in the overall economy, or at each firm, from the analysis. Industry groupings are based on SIC divisions. “Continuing firms only” only include firms, and people at firms, that are in the sample in both 1981 and 2013 (though the individuals at those firms are likely different).

Table A.4 – Complete Robustness Checks on Variance Decomposition - Part 2

	Total Var, 1981	Between-Firm Var, 1981	Total Var, 2013	Between-Firm Var, 2013	Total Var Increase	Frac Increase Between
Baseline sample	0.652	0.222	0.846	0.357	0.194	0.694
Age 20-29	0.556	0.178	0.62	0.241	0.063	0.993
Age 30-39	0.601	0.219	0.72	0.316	0.119	0.807
Age 40-49	0.631	0.255	0.791	0.348	0.16	0.583
Age 50-60	0.617	0.253	0.789	0.338	0.173	0.491
Continuing firms only	0.607	0.19	0.783	0.292	0.176	0.577
Midwest	0.632	0.221	0.736	0.281	0.104	0.585
Northeast	0.646	0.212	0.871	0.35	0.225	0.612
South	0.612	0.189	0.775	0.299	0.163	0.678
West	0.704	0.228	0.887	0.384	0.183	0.855
Size: 0-10	0.707	0.383	0.769	0.463	0.062	1.291
Size: 10-20	0.711	0.24	0.794	0.346	0.084	1.257
Size: 20-50	0.692	0.216	0.798	0.331	0.106	1.095
Size: 50-100	0.674	0.205	0.793	0.318	0.119	0.941
Size: 100-200	0.654	0.2	0.775	0.31	0.122	0.904
Size: 200-500	0.631	0.191	0.805	0.341	0.173	0.866
Size: 500-1k	0.599	0.17	0.831	0.362	0.232	0.827
Size: 1k-2k	0.582	0.161	0.838	0.364	0.256	0.795
Size: 2k-5k	0.595	0.174	0.867	0.382	0.272	0.764
Size: 5k-10k	0.602	0.176	0.9	0.395	0.298	0.737
Size: 10k+	0.552	0.164	0.873	0.348	0.32	0.577
Demean: county	0.611	0.181	0.8	0.311	0.189	0.687
Demean: state	0.63	0.2	0.828	0.339	0.198	0.701
Demean: census region	0.638	0.208	0.84	0.351	0.202	0.707
Demean: 2-digit SIC	0.554	0.125	0.75	0.261	0.196	0.697
Demean: 3-digit SIC	0.527	0.097	0.714	0.225	0.187	0.683
Demean: 4-digit SIC	0.517	0.088	0.705	0.216	0.187	0.684
Demean: gender	0.564	0.166	0.819	0.337	0.256	0.668
Demean: firm size category	0.611	0.182	0.838	0.349	0.226	0.738
Demean: person year of birth	0.568	0.186	0.695	0.26	0.127	0.578

Notes: See notes for Table A.3.

C.3 Further Robustness of Results on Earnings Inequality within and between Firms

Some robustness results are presented in Table II, and a much larger set of breakdowns are presented in Appendix Tables A.3 and A.4.

One concern not addressed in those tables 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.4 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 phenomenon (rather than a 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.5, we examine the sample of establishments in mega firms 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 workplaces in the United States.[7]

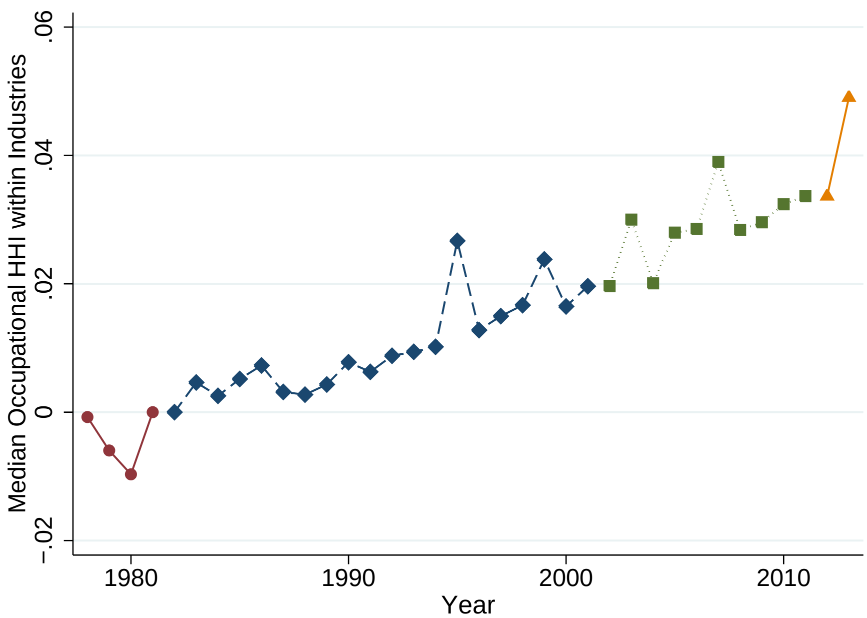
We also considered other robustness issues around health care, self-employment income, and business income. On health care, perhaps rising firm earnings inequality is offset by an increase in the generosity of firm health care insurance that, as a flat entitlement to all employees, provides a progressive compensation component. In fact, as Burkhauser and Simon (2010) show, *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 *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.[8]

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. Additionally, because this self-employment income is not generally connected to a particular firm, it is beyond the scope of this study on firms and inequality. To the extent that self-employment income is connected to particular firms—for example, in the increasing use of freelancers and independent contractors, as discussed by Weil (2014) and others—including that income would likely lead to higher estimates of sorting and segregation as previously diverse workforces shed all but a core group of likely similar employees.

[7] Barth et al. (2016) and Abowd et al. (2018) come to a similar conclusion.

[8] 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, conditional on total earnings, this part of health care is independent of the employee-firm match itself and does not influence our analysis.

Figure A.13 – Occupational Segregation Has Risen over Time



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.

We also unfortunately do not have data on business income; as noted by Smith et al. (2017), this income can be related to labor performed by firm owners and can be an important contributor to inequality at the very top of the income distribution. However, Smith et al. (2017) find that the firms owned by top business owners are often highly profitable rather than just large; indeed, these may be the same firms that pay other employees well, which would only amplify our results on between-firm inequality.

Overall, then, the basic result that the majority of increasing inequality is related to changes in firm average earnings seems to be broadly robust. One group that is a partial exception, as discussed in Section V.C.ii, is the top 1%. We also find that the rise in within-firm earnings inequality among the top 1% of earners 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 IV and V.

C.4 Further Exploration of Trends in within- and between-Firm Inequality: Outsourcing

One potential explanation for the results we find is increased outsourcing, as discussed in the main text. The rise in outsourcing is 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. (2016). Goldschmidt and Schmieder (2017) examine German data, finding clear evidence that a rise in outsourcing contributed to increasing inequality. This would lead firms to reorganize away from full-service production toward a more focused occupation structure. This is consistent with findings on the importance of outsourcing in rising inequality (Goldschmidt and Schmieder (2017)) and that occupations are increasingly concentrating within industries and firms (Kremer and Maskin (1996), Handwerker (2015)). We also find that industries are becoming increasingly concentrated by occupation; see Figure A.13. 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. On the other hand, any explanation involving outsourcing and occupational segregation would have to explain the differences in results between our paper and those from Germany (in Card et al. (2013)), where outsourcing and occupational segregation are also occurring.

D The Abowd, Kramarz and Margolis decomposition

D.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 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, Abowd et al. (2018)). Nevertheless, we are well aware of the limitations of the model, and incorporate them 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.

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.[9] However, as is well known, there is still a finite sample bias in estimates of $\text{var}_j(\psi^j)$ and $\text{var}_i(\theta^i)$ because of sampling error in the estimated worker and firm effects.

In addition, the estimate of the covariance term ($\text{cov}(\theta^i, \psi^j)$) 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; we discuss this further below. 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 A.5).

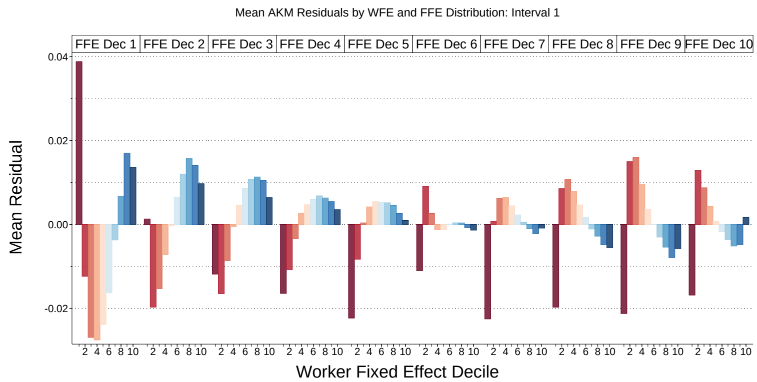
D.2 Model Fit

Table A.5 shows basic characteristics for the full sample of men as well as for observations of men 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 Appendix Tables A.8 and A.9.) Table A.5 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.

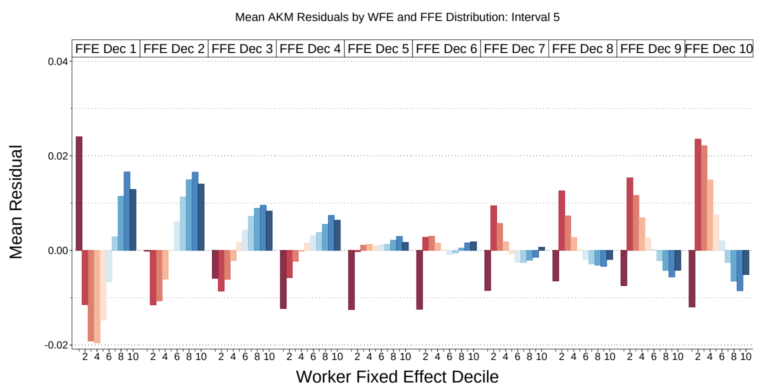
[9]The age-earnings gradient in SSA data flattens out around age 40. The worker effect is biased because it absorbs the time-invariant effect of age (i.e., age at start of the sample, which is effectively a cohort effect). Note that for the analysis of changes in the variance of worker effects over time, the normalization has no effect on the trend as long as the age distribution of the population and the return to age are roughly stable over time. The firm effects are not affected by the normalization. The covariance of worker and firm effects may be affected insofar as workers are sorted into firms by age.

FIGURE A.14 – Regression residuals by firm fixed effect decile

(A) 1980-1986



(B) 2007-2013

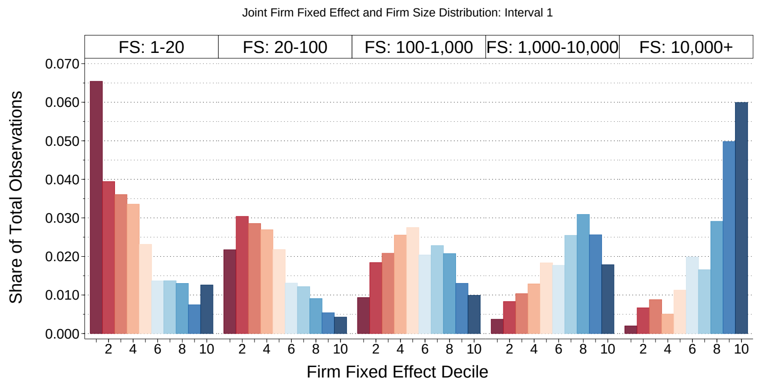


(C) Change from 1980-1986 to 2007-2013

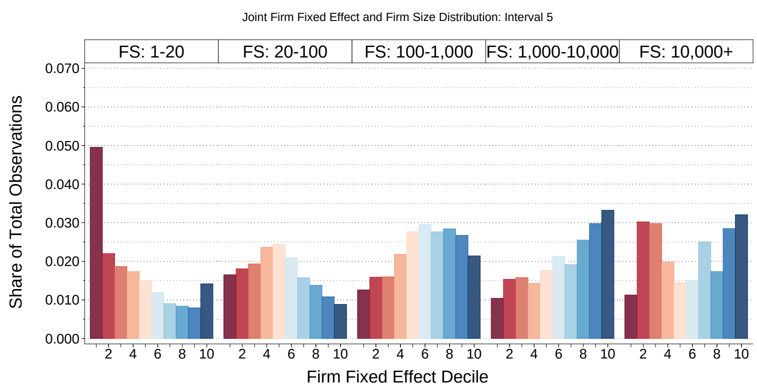
Notes: Calculations based on SSA data. Only men are included in these statistics. Only firms and individuals in firms with at least 20 employees are included. Only employed individuals aged 20 to 60 are included in all statistics, where "employed" is defined as earning the equivalent of 2013 minimum wage, adjusted for inflation with the PCE for 40 hours per week in 13 weeks. Individuals and firms in public administration or educational services are not included. Within each firm FE decile group, worker FE deciles are order from left to right from 1 to 10.

Figure A.15 – Distribution of workers among firm FE deciles, by firm size

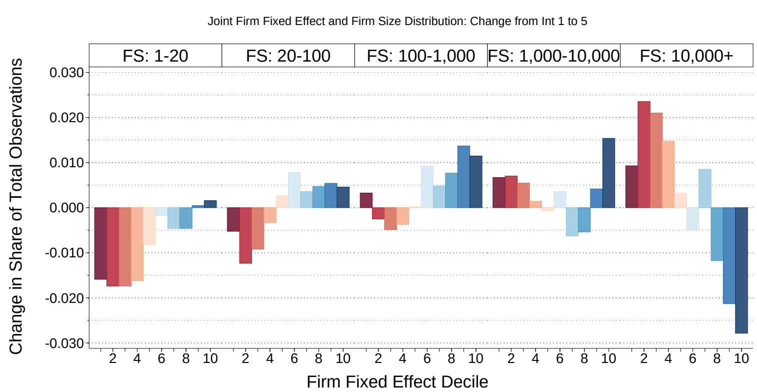
(a) 1980-1986



(b) 2007-2013



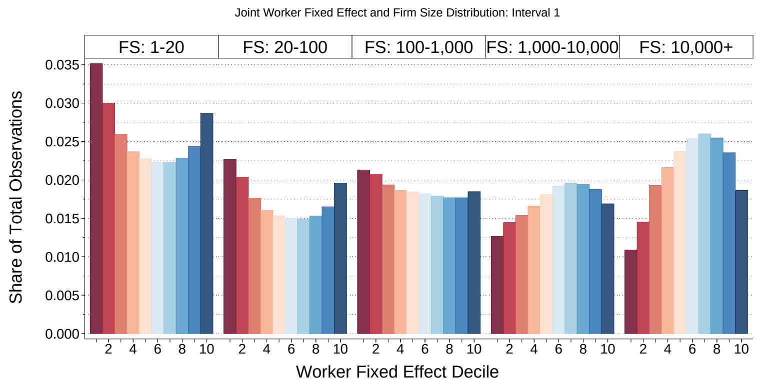
(c) Change from 1980-1986 to 2007-2013



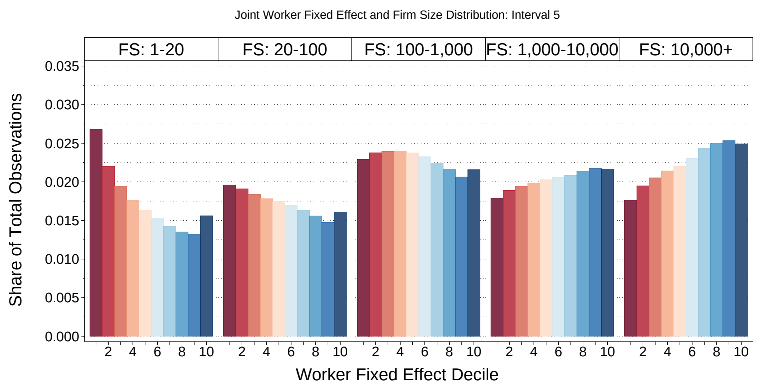
Notes: Calculations based on SSA data. Only men are included in these statistics. Only firms and individuals in firms with at least 20 employees are included. Only employed individuals aged 20 to 60 are included in all statistics, where "employed" is defined as earning the equivalent of 2013 minimum wage, adjusted for inflation with the PCE for 40 hours per week in 13 weeks. Individuals and firms in public administration or educational services are not included. Within each firm size group, firm FE deciles are order from left to right from 1 to 10.

Figure A.16 – Distribution of workers among worker FE deciles, by firm size

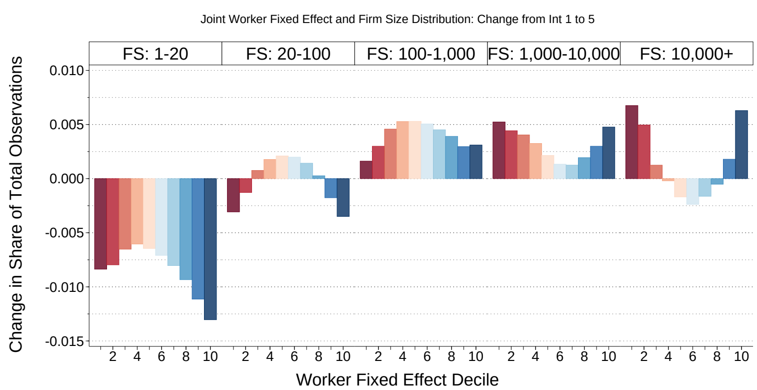
(a) 1980-1986



(b) 2007-2013



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



Notes: Calculations based on SSA data. Only men are included in these statistics. Only firms and individuals in firms with at least 20 employees are included. Only employed individuals aged 20 to 60 are included in all statistics, where "employed" is defined as earning the equivalent of 2013 minimum wage, adjusted for inflation with the PCE for 40 hours per week in 13 weeks. Individuals and firms in public administration or educational services are not included. Within each firm size group, worker FE deciles are order from left to right from 1 to 10.

Table A.5 – Summary Statistics for Overall Sample and Individuals in Largest Connected Set

	All employed men, age 20-60						Individuals in largest connected set					
				Log real annual earnings						Log real annual earnings		
7-year Interval	Number of worker/yr. obs. (1)	Number of workers (2)	Number of firms (3)	Mean (4)	Median (5)	Std. dev. (6)	Number of worker/yr. obs. (7)	Number of workers (8)	Number of firms (9)	Mean (10)	Median (11)	Std. dev. (12)
1980-1986	334,147,424	65,942,723	5,964,555	10.397	10.514	0.854	330,634,861 (98.9)	64,987,602 (98.6)	5,165,584 (86.6)	10.396 (100.0)	10.515 (100.0)	0.851 (99.6)
1987-1993	372,145,155	71,152,434	6,541,630	10.394	10.493	0.892	367,763,450 (98.8)	70,062,879 (98.5)	5,627,188 (86.0)	10.393 (100.0)	10.494 (100.0)	0.888 (99.5)
1994-2000	405,222,444	76,179,038	6,873,943	10.445	10.527	0.910	399,981,325 (98.7)	74,930,695 (98.4)	5,822,056 (84.7)	10.444 (100.0)	10.528 (100.0)	0.907 (99.7)
2001-2007	427,033,756	81,260,292	7,134,061	10.514	10.593	0.935	420,186,588 (98.4)	79,688,393 (98.1)	5,816,098 (81.5)	10.514 (100.0)	10.595 (100.0)	0.932 (99.8)
2007-2013	421,150,246	82,515,998	6,735,729	10.496	10.572	0.951	413,228,494 (98.1)	80,665,231 (97.8)	5,232,154 (77.7)	10.498 (100.0)	10.575 (100.0)	0.949 (99.8)
Change from first to last interval				0.099	0.057	0.096				0.102	0.060	0.099

Notes: Ratio of largest connected set to all observations in parentheses. Only men are included in these statistics. Individuals are included regardless of industry or firm size. Only employed individuals aged 20 to 60 are included in all statistics, where "employed" is defined as earning the equivalent 520 times the contemporaneous minimum wage.

Table A.6 – Estimation Results for AKM Model, Fit by Interval

		Interval 1 1980-1986 (1)	Interval 2 1987-1993 (2)	Interval 3 1994-2000 (3)	Interval 4 2001-2007 (4)	Interval 5 2007-2013 (5)
Sample	# Worker effects	64,987,602	70,062,879	74,930,695	79,688,393	80,665,231
Summary	# Firm effects	5,165,584	5,627,188	5,822,056	5,816,098	5,232,154
Statistics	Sample size	330,634,861	367,763,450	399,981,325	420,186,588	413,228,494
	sd(log(y))	0.851	0.888	0.907	0.932	0.949
Summary of	sd(WE)	0.587	0.624	0.657	0.677	0.693
AKM Parameter	sd(FE)	0.338	0.323	0.305	0.319	0.326
Estimates	sd(Xb)	0.239	0.261	0.282	0.249	0.243
	corr(WE,FE)	0.028	0.078	0.117	0.130	0.145
	corr(WE,Xb)	0.106	0.087	0.031	0.076	0.102
	corr(FE,Xb)	0.123	0.126	0.109	0.123	0.141
	rmse(residual)	0.431	0.427	0.421	0.428	0.411
	Adj R^2	0.743	0.768	0.784	0.789	0.812
Comparison	rmse(match residual)	0.365	0.363	0.354	0.360	0.346
Match Model	Adj R^2	0.816	0.833	0.848	0.851	0.867
	sd(match effect)	0.254	0.250	0.255	0.256	0.241

Notes: Only men are included in these statistics. Individuals are included regardless of industry or firm size. Only employed individuals aged 20 to 60 are included in all statistics, where "employed" is defined as earning the equivalent 520 times the 2013 PCE-deflated minimum wage.

Table A.6 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.[10] 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% in 1980-1986 to 81% 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 for a match effect reduces the RMSE and increases the adjusted R^2, by about the same amount

[10] 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.

each period, to $81-87\%$. 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.14b 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; because the dependent variable is in logs, this means that the predicted value and actual average generally differ by under 2%. 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.[11] Figure A.14c shows the change in mean residuals within cells over time. The changes are of opposite signs of the deviations in A.14b, 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.

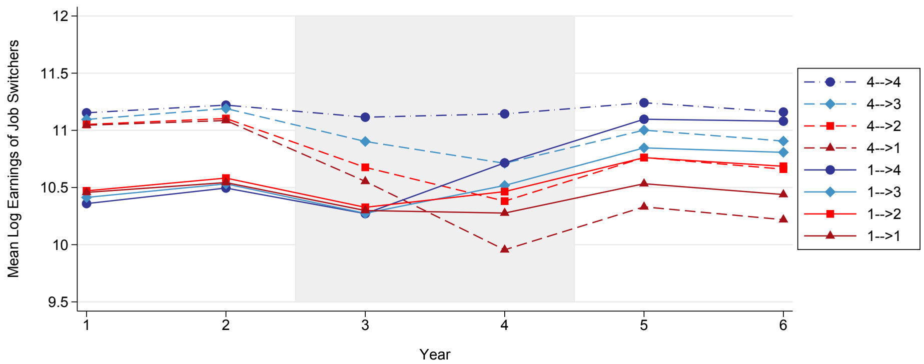
Another diagnostic assesses the ability of the model to explain earnings changes at job changes. If the model is correctly specified, on average workers changing from one firm to another should experience earnings changes corresponding to the estimated firm effects. To implement this comparison, we used 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 V and A.17, for firm classes based on firm fixed effects and firm average wages, respectively. The results are discussed in the main text in Section IV. Overall, we conclude that despite the fact we are modeling information on annual earnings rather than

[11] Not surprisingly given the presence of labor supply effects, the mean residuals in Figure A.14 are on average larger than those shown CHK (Figure VI).

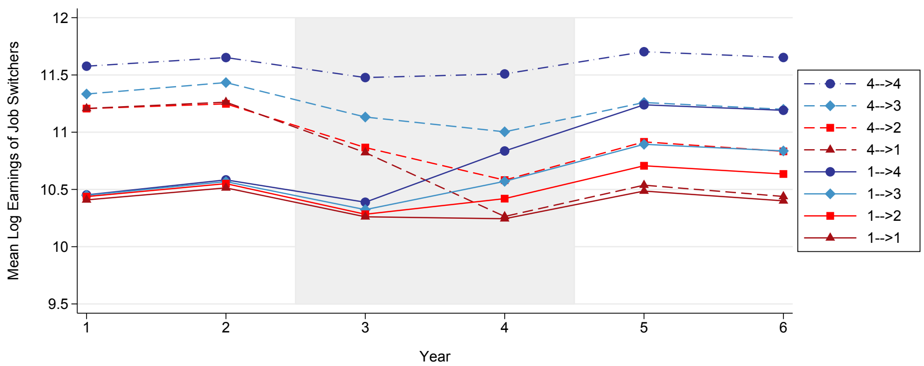
daily or hourly wages, our model delivers a good approximation of the underlying earnings process.

Figure A.17 – 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. Only men are included in these statistics. Men are included regardless of industry or firm size. Only employed individuals aged 20 to 60 are included in all statistics, where "employed" is defined as earning the equivalent 520 times the contemporaneous minimum wage. For an explanation of the methodology see Section IV.B and Appendix D.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.

Several other robustness checks are shown in Table A.7, all of which show fairly similar results. The second column shows that the main results are similar when restricting the sample to women only, rather than men only. For the third column, we construct a "sandwich sample" to include only observations on either side of a job move: that is, we drop an observation for an individual in year t if that individual's primary EIN in year t is different from their EIN in years $t-1$ or $t+1$. This sandwich sample helps ensure that our results are not driven by changes in labor supply caused by the process of switching jobs. Finally, Lachowska et al. (2017) find that the use of hourly wage rather than annual earnings can change estimated variance components. Although we do not observe hours, we restrict our sample to those with at least the equivalent of 2000 hours at the 2013 minimum wage in the last column of Table A.7; based on an analysis of CPS data, this ensures that the vast majority of the sample works full time, mostly eliminating variation in hours. The similarity of the main results to the results of the specifications with labor supply restrictions suggest that our results are being driven more by wage differences than by differences in hours or weeks of work.

Finally, Tables A.8 and A.9 show more details on our main results for the women-only sample. Women see somewhat less rising between-firm variance than men. However, for women as well as men, rising between-firm variance still accounts for most of rising inequality; rising dispersion of individual fixed effects accounts for most of rising variance in the basic decomposition; and, in the detailed decomposition, sorting and segregation dominate the between-firm component.

Table A.12 shows detailed descriptive statistics on the characteristics of job movers. The first row shows the average number of moves per interval. The table shows a slight decline in the average number of job moves per worker. However, we suspect that this is due to the effect of the Great Recession in 2007-2013 as prior to this interval the series shows no trend. We also include average age, earnings, worker fixed effect, firm size, and firm fixed effect of job movers versus job stayers. As expected, job movers tend to be younger, come from smaller firms, and have lower average earnings. Job movers also on average have lower firm fixed effects than job stayers. In addition, if we disaggregate job movers by the number of switches per interval, we find a negative, monotonic relationship between the number of switches and the average firm fixed effect. This in turn implies a positive relationship between match duration and firm fixed effects. What is important for our purposes is that the average years in the sample of movers and stayers is similar and has remained similar over time, suggesting that differential selection into the sample is unlikely to affect our estimates.

Table A.13 reports the correlation of estimated firm effects across intervals. The estimated firm fixed effects are significantly correlated across intervals and this correlation drops only slightly as the time period is extended to a three interval difference. In particular, the firm fixed effects of large firms are highly correlated over time.

One further concern with the AKM estimator is limited mobility bias. As discussed by Andrews et al. (2008), because identification of fixed effects is based on flows of workers between firms, the small numbers of workers moving in and out of some firms–and the small number of firms that employ most workers–leads to imprecise estimates of fixed effects. This imprecision in turn causes variance estimates to be biased. To address this, we first follow Andrews et al. (2012) in estimating the model with a 10% sample (not shown here), which should exacerbate the bias. As expected, variances of fixed effects are higher and covariances are lower, but

Table A.7 – Sensitivity of AKM Results

	Main Sample: Male				Main Sample: Female				Sandwich Sample: Male				"Full-time" Sample: Male			
	Int 1 (1)	Int 5 (2)	Change (3)	Share (4)	Int 1 (5)	Int 5 (6)	Change (7)	Share (8)	Int 1 (9)	Int 5 (10)	Change (11)	Share (12)	Int 1 (13)	Int 5 (14)	Change (15)	Share (16)
Total Var	0.71	0.92	0.22	-	0.53	0.75	0.22	-	0.39	0.61	0.22	-	0.36	0.53	0.17	-
Between Firm Var	0.21	0.37	0.16	74.2	0.17	0.28	0.12	53.5	0.12	0.26	0.14	63.7	0.09	0.19	0.10	57.9
Var(m_wfe)	0.05	0.12	0.07	30.9	0.04	0.09	0.05	21.4	0.06	0.12	0.07	30.5	0.04	0.09	0.06	34.9
Var(m_ffe)	0.08	0.08	0.00	-1.6	0.07	0.07	0.00	-0.4	0.08	0.10	0.02	7.4	0.04	0.03	0.00	-2.2
Var(m_xb)	0.01	0.01	0.00	0.6	0.00	0.01	0.00	1.5	0.00	0.00	0.00	0.0	0.00	0.00	0.00	-0.2
2Cov(m_wfe,m_ffe)	0.03	0.11	0.08	34.8	0.05	0.09	0.04	18.2	-0.03	0.01	0.05	21.2	0.01	0.05	0.04	23.3
2Cov(m_wfe,m_xb)	0.01	0.03	0.02	7.3	0.00	0.02	0.01	6.6	0.00	0.01	0.01	2.7	0.00	0.01	0.00	2.0
2Cov(m_ffe,m_xb)	0.02	0.03	0.00	2.2	0.01	0.02	0.01	6.1	0.01	0.01	0.00	1.8	0.00	0.00	0.00	0.1
Within Firm Var	0.50	0.55	0.06	25.8	0.36	0.47	0.10	46.5	0.27	0.35	0.08	36.3	0.27	0.34	0.07	42.1
Var(diff_wfe)	0.28	0.36	0.08	36.7	0.21	0.30	0.09	42.4	0.19	0.29	0.10	45.0	0.17	0.26	0.08	51.0
Var(diff_xb)	0.05	0.05	0.00	1.2	0.03	0.04	0.01	5.3	0.02	0.02	0.00	-0.1	0.02	0.02	0.00	0.3
Var(diff_r)	0.15	0.14	-0.02	-8.2	0.13	0.12	-0.01	-4.8	0.05	0.03	-0.02	-8.5	0.08	0.07	-0.01	-5.9
2Cov(diff_wfe,diff_xb)	0.02	0.01	-0.01	-3.2	-0.01	0.00	0.01	3.7	0.01	0.01	0.00	0.0	0.00	0.00	-0.01	-3.4
Segregation index	0.16	0.25	0.09		0.16	0.22	0.07		0.23	0.30	0.07		0.18	0.27	0.09	
Corr(wfe,ffe)	0.10	0.28	0.18		0.18	0.27	0.09		-0.12	0.03	0.15		0.06	0.23	0.17	
N (millions)	221.62	302.77	81.15		159.45	268.25	108.80		127.45	152.84	25.39		188.83	253.38	64.55	
Adj R_2	0.743	0.812	0.07		0.694	0.796	0.10		0.835	0.882	0.05		0.753	0.840	0.09	
Adj R_2 - match model	0.816	0.867	0.05		0.772	0.857	0.09		0.845	0.920	0.08		0.791	0.875	0.08	

Notes: Variance and correlation of fixed effects estimated by AKM model as explained in Section IV. Only firms and individuals in firms with at least 20 employees are included. Only employed individuals aged 20 to 60 are included in all statistics. Individuals and firms in public administration or educational services are not included, and only those aged 20 to 60 are included. The first column includes the usual sample restrictions: men only, and employed is defined as earning the equivalent of 2013 minimum wage, adjusted for inflation with the PCE for 40 hours per week in 13 weeks. The second column includes women only. The third column (the "sandwich sample") drops observations for individuals in years that are not preceded and succeeded by observations with the same primary EIN. The fourth column restricts the analysis to "full-time" workers, defined as earning the equivalent of the 2013 minimum wage for 2000 hours.

TABLE A.8 – Basic Decomposition of the Rise in Inequality of Annual Earnings: Women Only

		Interval 1 (1980-1986)		Interval 2 (1987-1993)		Interval 3 (1994-2000)		Interval 4 (2001-2007)		Interval 5 (2007-2013)		Change from 1 to 5	
		Comp (1)	Share (2)	Comp (3)	Share (4)	Comp (5)	Share (6)	Comp (7)	Share (8)	Comp (9)	Share (10)	Comp (11)	Share (12)
Total Variance	Var(y)	0.530	-	0.594	-	0.657	-	0.719	-	0.751	-	0.222	-
Components of variance	Var(WFE)	0.245	46.2	0.288	48.5	0.330	50.3	0.365	50.7	0.386	51.4	0.142	63.8
	Var(FFE)	0.070	13.1	0.062	10.4	0.058	8.8	0.067	9.3	0.069	9.1	-0.001	-0.4
	Var(Xb)	0.030	5.6	0.035	5.9	0.055	8.4	0.046	6.4	0.045	6.0	0.015	6.9
	Var(residual)	0.135	25.5	0.137	23.0	0.139	21.2	0.140	19.4	0.124	16.5	-0.011	-4.8
	2*Cov(WFE,FFE)	0.047	8.8	0.055	9.3	0.062	9.5	0.074	10.3	0.087	11.6	0.040	18.2
	2*Cov(WFE,Xb)	-0.003	-0.6	0.007	1.2	-0.001	-0.2	0.011	1.5	0.020	2.6	0.023	10.4
	2*Cov(FFE,Xb)	0.007	1.3	0.010	1.6	0.014	2.1	0.017	2.3	0.020	2.7	0.014	6.1
Sum of firm components	Cov(y,FFE)	0.096	18.2	0.094	15.8	0.096	14.6	0.112	15.6	0.122	16.3	0.026	11.8
Counterfactuals	1.) No rise in Corr(WFE,FFE)	0.530		0.586	98.7	0.644	98.1	0.701	97.5	0.723	96.2	0.193	87.1
	2.) No fall in Var(FFE)	0.530		0.606	102.0	0.676	103.0	0.724	100.7	0.753	100.2	0.224	100.8
	3.) Both 1 and 2	0.530		0.597	100.6	0.662	100.8	0.705	98.0	0.724	96.4	0.195	87.7

Notes: Results for women only; otherwise, see notes for Table III.

Table A.9 – Detailed Decomposition of the Rise in Earnings Inequality between and within Firms: Women Only

		Interval 1 (1980-1986)		Interval 2 (1987-1993)		Interval 3 (1994-2000)		Interval 4 (2001-2007)		Interval 5 (2007-2013)		Change from 1 to 5	
		Comp (1)	Share (2)	Comp (3)	Share (4)	Comp (5)	Share (6)	Comp (7)	Share (8)	Comp (9)	Share (10)	Comp (11)	Share (12)
Total var	Var(y)	0.530	-	0.594	-	0.657	-	0.719	-	0.751	-	0.222	-
Between-firm var	Var(m_y)	0.166	31.4	0.190	32.0	0.216	32.9	0.248	34.6	0.285	37.9	0.119	53.5
	Var(m_WFE)	0.038	7.2	0.053	8.9	0.065	9.9	0.073	10.1	0.086	11.4	0.047	21.4
	Var(m_FFE)	0.070	13.1	0.062	10.4	0.058	8.8	0.067	9.3	0.069	9.1	-0.001	-0.4
	Var(m_Xb)	0.003	0.5	0.004	0.6	0.006	0.9	0.005	0.7	0.006	0.8	0.003	1.5
	2Cov(m_WFE,m_FFE)	0.047	8.8	0.055	9.3	0.062	9.5	0.074	10.3	0.087	11.6	0.040	18.2
	2Cov(m_WFE,m_Xb)	0.003	0.5	0.007	1.2	0.011	1.7	0.013	1.8	0.017	2.3	0.015	6.6
	2Cov(m_FFE,m_Xb)	0.007	1.3	0.010	1.6	0.014	2.1	0.017	2.3	0.020	2.7	0.014	6.1
Within-firm var	Var(diff_y)	0.363	68.6	0.404	68.0	0.440	67.1	0.470	65.4	0.466	62.1	0.103	46.5
	Var(diff_WFE)	0.207	39.0	0.235	39.6	0.265	40.3	0.292	40.6	0.301	40.0	0.094	42.4
	Var(diff_Xb)	0.027	5.2	0.032	5.3	0.049	7.5	0.041	5.7	0.039	5.2	0.012	5.3
	Var(diff_r)	0.135	25.5	0.137	23.0	0.139	21.2	0.140	19.4	0.124	16.5	-0.011	-4.8
	2Cov(diff_WFE,diff_Xb)	-0.006	-1.1	0.000	0.0	-0.012	-1.9	-0.002	-0.2	0.003	0.3	0.008	3.7
	2Cov(diff_WFE,diff_r)	0.000	0.1	0.000	0.1	0.000	0.0	0.000	0.0	0.000	0.0	-0.001	-0.3
	2Cov(diff_Xb,diff_r)	0.000	-0.1	0.000	0.0	0.000	-0.1	0.000	0.0	0.000	0.0	0.000	0.1
Segregation Index	Var(m_WFE)/Var(WFE)	0.156		0.183		0.198		0.199		0.221		0.066	
	N (millions)	159.45		206.46		239.43		263.02		268.25		108.80	

Notes: Results for women only; otherwise, see notes for Table IV.

TABLE A.10 – Basic Decomposition of the Rise in Inequality: All Firm Sizes, All Industries

		Interval 1 (1980-1986)		Interval 2 (1987-1993)		Interval 3 (1994-2000)		Interval 4 (2001-2007)		Interval 5 (2007-2013)		Change from 1 to 5	
		Comp (1)	Share (2)	Comp (3)	Share (4)	Comp (5)	Share (6)	Comp (7)	Share (8)	Comp (9)	Share (10)	Comp (11)	Share (12)
Total Variance	Var(y)	0.724	-	0.788	-	0.822	-	0.869	-	0.901	-	0.177	-
Components of variance	Var(WFE)	0.345	47.6	0.390	49.4	0.431	52.4	0.458	52.7	0.480	53.2	0.135	76.1
	Var(FFE)	0.114	15.8	0.104	13.2	0.093	11.3	0.102	11.7	0.106	11.8	-0.008	-4.5
	Var(Xb)	0.057	7.9	0.068	8.6	0.080	9.7	0.062	7.1	0.059	6.5	0.002	1.0
	Var(residual)	0.147	20.3	0.145	18.4	0.141	17.2	0.146	16.8	0.134	14.9	-0.013	-7.2
	2*Cov(WFE,FFE)	0.011	1.5	0.031	4.0	0.047	5.7	0.056	6.5	0.065	7.3	0.054	30.6
	2*Cov(WFE,Xb)	0.030	4.1	0.028	3.6	0.011	1.4	0.026	2.9	0.034	3.8	0.005	2.6
	2*Cov(FFE,Xb)	0.020	2.7	0.021	2.7	0.019	2.3	0.020	2.2	0.022	2.5	0.002	1.4
Sum of firm components	Cov(y,FFE)	0.130	17.9	0.131	16.6	0.126	15.3	0.140	16.1	0.150	16.7	0.020	11.4
Counterfactuals	1.) No rise in Corr(WFE,FFE)	0.724		0.768		0.787		0.825		0.848		0.125	70.3
	2.) No fall in Var(FFE)	0.724		0.801		0.851		0.886		0.912		0.189	106.4
	3.) Both 1 and 2	0.724		0.780		0.811		0.840		0.858		0.134	75.6

Notes: Only men are included in these statistics. Individuals are included regardless of industry or firm size. Only employed individuals aged 20 to 60 are included in all statistics, where "employed" is defined as earning the equivalent of 520 times the 2013 PCE deflated minimum wage.

TABLE A.11 – Detailed Decomposition of the Rise in Inequality between and within Firms: All Firms, All Industries

		Interval 1 (1980-1986)		Interval 2 (1987-1993)		Interval 3 (1994-2000)		Interval 4 (2001-2007)		Interval 5 (2007-2013)		Change from 1 to 5	
		Comp (1)	Share (2)	Comp (3)	Share (4)	Comp (5)	Share (6)	Comp (7)	Share (8)	Comp (9)	Share (10)	Comp (11)	Share (12)
Total var	Var(y)	0.724	-	0.788	-	0.822	-	0.869	-	0.901	-	0.177	-
Between-firm var	Var(m_y)	0.250	34.5	0.287	36.4	0.307	37.3	0.332	38.2	0.369	41.0	0.120	67.4
	Var(m_wfe)	0.081	11.2	0.100	12.7	0.117	14.3	0.126	14.5	0.142	15.8	0.062	34.7
	Var(m_ffe)	0.114	15.8	0.104	13.2	0.093	11.3	0.102	11.7	0.106	11.8	-0.008	-4.5
	Var(m_xb)	0.009	1.2	0.011	1.4	0.011	1.3	0.009	1.0	0.009	0.9	0.000	-0.1
	2Cov(m_wfe,m_ffe)	0.011	1.5	0.031	4.0	0.047	5.7	0.056	6.5	0.065	7.3	0.054	30.6
	2Cov(m_wfe,m_xb)	0.015	2.0	0.020	2.5	0.020	2.4	0.020	2.3	0.024	2.7	0.010	5.4
	2Cov(m_ffe,m_xb)	0.020	2.7	0.021	2.7	0.019	2.3	0.020	2.2	0.022	2.5	0.002	1.4
Within-firm var	Var(diff_y)	0.474	65.5	0.501	63.6	0.515	62.7	0.537	61.8	0.532	59.0	0.058	32.6
	Var(diff_wfe)	0.264	36.5	0.290	36.8	0.314	38.1	0.333	38.2	0.337	37.4	0.073	41.4
	Var(diff_xb)	0.048	6.7	0.057	7.3	0.069	8.3	0.053	6.1	0.050	5.6	0.002	1.1
	Var(diff_r)	0.147	20.3	0.145	18.4	0.141	17.2	0.146	16.8	0.134	14.9	-0.013	-7.2
	2Cov(diff_wfe,diff_xb)	0.015	2.1	0.009	1.1	-0.008	-1.0	0.006	0.6	0.010	1.1	-0.005	-2.8
	2Cov(diff_wfe,diff_r)	0.000	0.0	0.000	0.0	0.000	0.0	0.000	0.0	0.000	0.0	0.000	0.0
	2Cov(diff_xb,diff_r)	0.000	0.0	0.000	0.0	0.000	0.0	0.000	0.0	0.000	0.0	0.000	0.0
Segregation index	Var(m_wfe)/Var(wfe)	0.234		0.256		0.272		0.274		0.297		0.062	
	N (millions)	331		368		400		420		413		82	

Notes: Only men are included in these statistics. Individuals are included regardless of industry or firm size. Only employed individuals aged 20 to 60 are included in all statistics, where "employed" is defined as earning the equivalent of 520 times the 2013 PCE deflated minimum wage.

Table A.12 – Descriptive Statistics for Job Movers versus Job Stayers

	Type of worker	Interval 1 (1980-1986) (1)	Interval 2 (1987-1993) (2)	Interval 3 (1994-2000) (3)	Interval 4 (2001-2007) (4)	Interval 5 (2007-2013) (5)	Change from 1 to 5 (6)	Average (1980-2013) (7)
Mean number of job moves	*All workers*	1.02	1.03	1.14	1.06	0.91	-0.11	1.03
	Job mover	1.82	1.89	1.97	1.91	1.79	-0.03	1.88
	Job stayer	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	One year in sample	-	-	-	-	-	-	-
Percent of observations	*Job mover*	55.97	54.49	58.03	55.55	50.68	-5.29	54.89
	Job stayer	34.95	37.16	33.52	35.62	39.37	4.41	36.17
	One year in sample	9.08	8.35	8.45	8.83	9.95	0.88	8.94
Mean years in sample	*Job mover*	5.55	5.67	5.80	5.74	5.62	0.07	5.68
	Job stayer	5.43	5.67	5.69	5.64	5.52	0.10	5.59
	One year in sample	1.00	1.00	1.00	1.00	1.00	0.00	1.00
Mean age	*Job mover*	31.34	31.53	33.00	33.86	34.07	2.73	32.85
	Job stayer	36.89	37.54	39.12	40.04	40.40	3.50	38.91
	One year in sample	33.37	34.05	34.17	34.94	35.87	2.50	34.54
Mean firm size (thousands)	*Job mover*	294.1	271.9	232.0	309.9	369.9	75.84	296.64
	Job stayer	723.3	675.3	527.0	569.7	669.4	-53.90	628.56
	One year in sample	373.6	282.8	257.6	352.2	424.8	51.15	338.60
Mean log earnings	*Job mover*	10.08	10.00	10.12	10.16	10.15	0.07	10.10
	Job stayer	10.59	10.60	10.65	10.71	10.67	0.08	10.65
	One year in sample	9.37	9.26	9.34	9.34	9.40	0.03	9.34
Mean worker FE	*Job mover*	10.39	10.23	9.67	10.64	10.53	0.14	10.29
	Job stayer	10.73	10.63	10.07	11.05	10.93	0.20	10.69
	One year in sample	9.95	9.79	9.22	10.13	10.09	0.14	9.84
Mean firm FE	*Job mover*	-0.05	0.08	0.67	-0.21	-0.09	-0.04	0.08
	Job stayer	0.08	0.19	0.73	-0.14	-0.03	-0.11	0.17
	One year in sample	-0.15	-0.02	0.56	-0.32	-0.21	-0.07	-0.03

Notes: "One year in sample" refers to workers workers who are in the sample only a single year and hence cannot be classified as movers or stayers. Statistics are computed at the worker-interval level. For each individual in each interval we compute number of years in the sample, number of job switches, average age, average earnings, average firm size, worker fixed effect, and average firm fixed effect. Then we compute descriptive statistics based on the worker-level variables. Only men are included in these statistics. Individuals are included regardless of industry or firm size. Only employed individuals aged 20 to 60 are included in all statistics, where "employed" is defined as earning the equivalent of 520 times the contemporaneous minimum wage.

changes over time are similar. As a further test, in Table A.14, we compare estimates for samples of firms with different numbers of connections to other firms. We measure connections as the total number of a given firm's employees that worked for other firms in a given seven-year interval. Since we only use a worker main employer in a given year for the analysis, this captures worker flows between firms, which is the relevant measure for assessing the precision of AKM estimates.[12] As expected, the level of variances of fixed effects is higher and covariances

[12] For each employer i of firm j in a given seven-year interval p, we first calculate the number of other employers k_{ij} the worker worked at. For each firm, we then sum up to obtain the number of workers that had other employers to obtain our measure of connectedness $K_j = \sum_{i \in j}(k_{ij} - 1)$. For example, if a firm has a 100 employees, 50 of which worked at one other firm in a seven-year interval, then $k_{ij} = 2$ for all 50 of these workers, and $K_j = 50$. This number is the same whether these workers all came from

Table A.13 – Correlation of Firm Fixed Effects across Time

	All firms (1)	Large firms (2)	Medium firms (3)	Small firms (4)
A. Correlation across one interval				
1980-1986 to 1987-1993	0.67	0.92	0.75	0.39
1987-1993 to 1994-2000	0.67	0.88	0.80	0.40
1994-2000 to 2001-2007	0.66	0.84	0.80	0.40
2001-2007 to 2007-2013	0.72	0.90	0.87	0.46
B. Correlation across two intervals				
1980-1986 to 1994-2000	0.60	0.80	0.67	0.33
1987-1993 to 2001-2007	0.60	0.72	0.74	0.36
1994-2000 to 2007-2013	0.64	0.84	0.78	0.36
C. Correlation across three intervals				
1980-1986 to 2001-2007	0.57	0.68	0.63	0.32
1987-1993 to 2007-2013	0.61	0.77	0.74	0.33

Notes: Fixed effects are weighted by the average number of observations across intervals. Large firms have more than 10,000 employees, medium firms have 101 to 10,000 employees, small firms have 1 to 100 employees. Employment is counted in terms of males aged 20-60 years. Only men are included in these statistics. Individuals are included regardless of industry or firm size. Only employed individuals aged 20 to 60 are included in all statistics, where "employed" is defined as earning the equivalent of 520 times the contemporaneous minimum wage.

are lower for groups of firms with fewer connections. The trends in variances and covariances of fixed effects estimates based on firms with smaller numbers of connections are somewhat weaker, but the overall direction is again the same.

the same or from different employers.

Table A.14 – Variance and Correlation of Fixed Effects by Firms' Connectedness to Other Firms Via Worker Flows

Number of Employees Connected to Other Firms	Time Period	Percent of Obs.	Std(WFE)	Std(FFE)	Corr(WFE,FFE)
1 to 5	Interval 1: 1980-1986	0.1	0.947	0.659	-0.244
	Interval 5: 2007-2013	0.1	1.113	0.634	-0.202
	Change	0.0	0.166	-0.025	0.041
6 to 10	Interval 1: 1980-1986	0.2	0.804	0.472	-0.042
	Interval 5: 2007-2013	0.2	0.946	0.441	0.013
	Change	0.1	0.142	-0.032	0.055
11 to 20	Interval 1: 1980-1986	0.5	0.735	0.371	0.041
	Interval 5: 2007-2013	0.7	0.845	0.365	0.123
	Change	0.2	0.109	-0.007	0.082
21 to 50	Interval 1: 1980-1986	2.8	0.674	0.300	0.085
	Interval 5: 2007-2013	3.5	0.751	0.303	0.196
	Change	0.7	0.078	0.002	0.111
Over 50	Interval 1: 1980-1986	96.4	0.569	0.286	0.105
	Interval 5: 2007-2013	95.4	0.684	0.281	0.287
	Change	-1.1	0.115	-0.005	0.182

Note: Standard deviation and correlation of fixed effects estimated by AKM model as explained in Section IV. Only men are included in these statistics. Only firms and individuals in firms with at least 20 employees are included. Only employed individuals aged 20 to 60 are included in all statistics, where "employed" is defined as earning the equivalent of 2013 minimum wage, adjusted for inflation with the PCE for 40 hours per week in 13 weeks. Individuals and firms in public administration or educational services are not included. "Number of Employees Connected to Other Firms" is defined as the total number of a given firm's employees that worked for another firm within a given seven-year interval. As in the remainder of the AKM analysis, this statistic relies only on workers' main employer in a given year, so this measures transitions between firms.

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

Tables III and IV 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 A.18 display 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 A.18a. 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 A.18b show these patterns have changed substantially over time. Figure A.18c shows the net change of density of the two distributions at corresponding deciles.[13] 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.

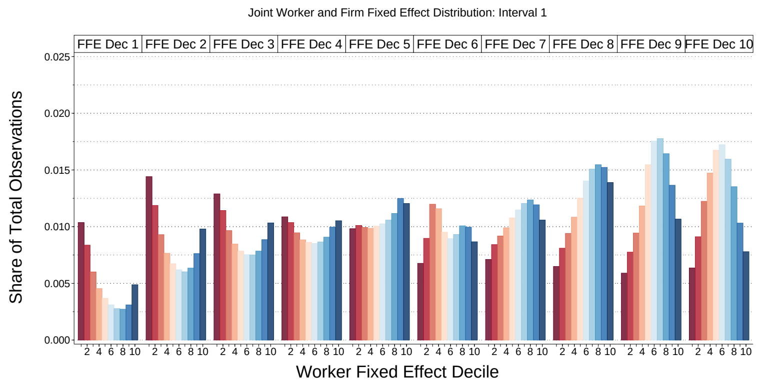
The figures confirm the evidence from the variance decomposition that sorting has increased, and show which workers and firms appear most affected. A striking finding is that the incidence and composition of workers at high-wage firms has been changing substantially. Since high-wage firms are likely 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.15 and A.16 for three firm size groups (firms with number of workers in range $1-100$, $101-9999$, and $10,000+$).

From Figure A.15 it is clear that on average, high-wage firms tend to be larger. However, over time, Figure A.15c shows that large employers have experienced a substantial shift out of high-wage firms to middle and lower-wage firms. Figure A.16 shows that among larger firms, the decline was accompanied by an *increase* in the incidence of high wage workers at larger

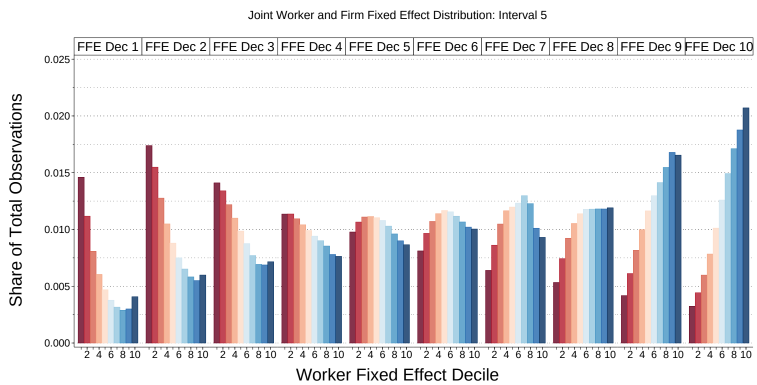
[13] Note that the definition of the deciles differ between the two time periods. Yet, since the distribution of firm effects has changed little, the deciles of firm effects are roughly comparable over time.

Figure A.18 – Distribution of Workers among Deciles of Worker and Firm Fixed Effects

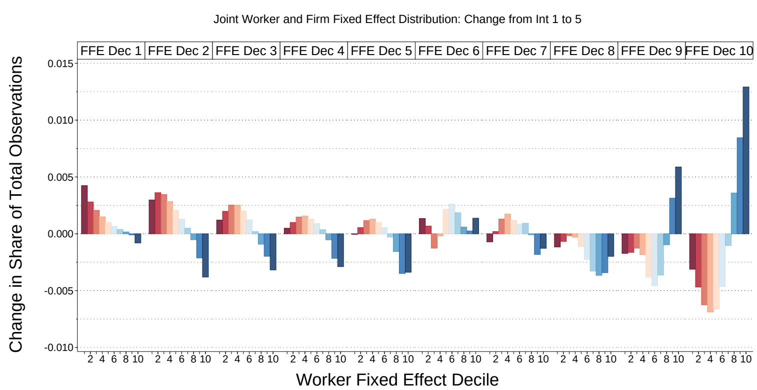
(a) 1980-1986



(b) 2007-2013



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



Notes: Calculations based on SSA data. Only men are included in these statistics. Only firms and individuals in firms with at least 20 employees are included. Only employed individuals aged 20 to 60 are included in all statistics, where "employed" is defined as earning the equivalent of 2013 minimum wage, adjusted for inflation with the PCE for 40 hours per week in 13 weeks. Individuals and firms in public administration or educational services are not included. Firm and worker fixed effects from our AKM estimation are 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 ordered from left to right from 1 to 10.

firms. In addition, mega firms saw a reduction in 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 can be produced upon request. 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 A.18 – 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 IV 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 also examined to what extent our main findings in Table IV 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 A.19 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. Each industry generally sees strong increases in correlations beginning in the early 1980s that are slowing down over time; an exception is education and public administration, in which the correlation has been flat (those sectors are dropped from other analyses that aggregate all industries). Table A.15 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.

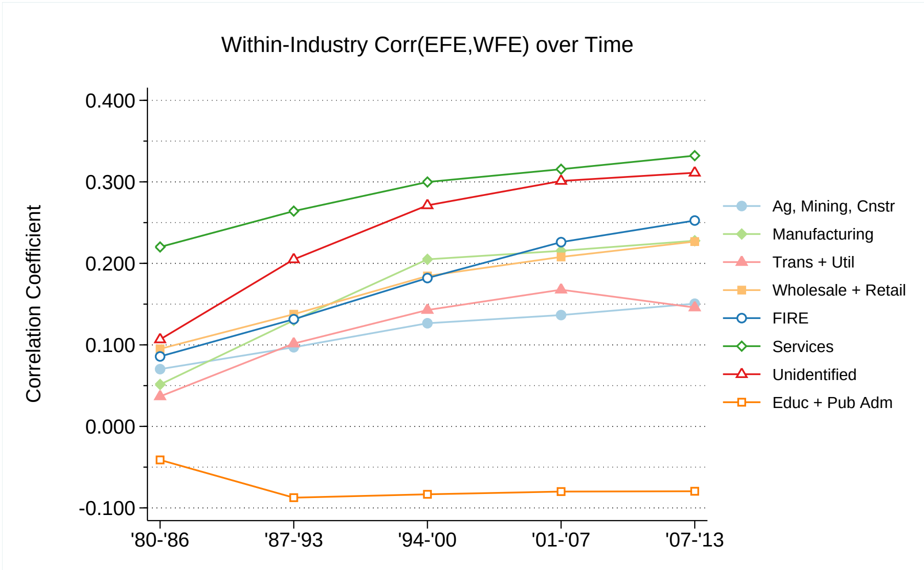
We have also examined the pattern of marginal distributions of firm and worker effects by one-digit industry. These figures can also be produced upon request. The results show that the large decline in the incidence of employment in higher wage deciles tends to be concentrated in manufacturing. Employment 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 firm with unknown industry affiliation. These are likely to be disproportionately new employers, which might be likely to have low firm fixed effects.

Table A.15 – Variation in the Distribution and Correlation of Worker and Firm Fixed Effects by Major Industry from 1980 to 2013

	Interval 1: 1980-1986				Interval 5: 2007-2013				Change from Interval 1 to 5			
	Emp. Share (1)	Var(WFE) (2)	Var(FFE) (3)	Corr(WFE,FFE) (4)	Emp. Share (5)	Var(WFE) (6)	Var(FFE) (7)	Corr(WFE,FFE) (8)	Emp. Share (9)	Var(WFE) (10)	Var(FFE) (11)	Corr(WFE,FFE) (12)
Agriculture, Forestry, & Fishing	2.03	0.201	0.032	0.135	1.35	0.312	0.047	0.165	-0.67	0.111	0.015	0.030
Mining	1.97	0.304	0.071	0.005	0.72	0.398	0.047	0.200	-1.25	0.094	-0.025	0.195
Construction	6.33	0.401	0.053	0.100	4.90	0.414	0.035	0.102	-1.42	0.013	-0.018	0.003
Manufacturing	31.43	0.279	0.068	0.051	13.00	0.412	0.052	0.228	-18.43	0.133	-0.016	0.176
Transportation & Public Utilities	9.35	0.244	0.063	0.037	6.54	0.342	0.069	0.146	-2.81	0.099	0.006	0.109
Wholesale Trade	5.09	0.363	0.045	0.051	4.62	0.434	0.045	0.186	-0.46	0.071	0.000	0.135
Retail Trade	8.82	0.389	0.036	0.097	8.74	0.429	0.033	0.164	-0.08	0.040	-0.004	0.067
Finance, Insurance, & Real Estate	5.57	0.427	0.075	0.086	5.52	0.636	0.077	0.253	-0.06	0.209	0.002	0.167
Services	14.04	0.383	0.094	0.220	23.46	0.524	0.082	0.332	9.42	0.140	-0.012	0.112
Education	2.54	0.400	0.059	-0.094	4.43	0.495	0.042	-0.104	1.89	0.095	-0.017	-0.011
Public Administration	9.87	0.215	0.075	0.057	8.69	0.303	0.041	0.024	-1.18	0.088	-0.034	-0.033
Unidentified	2.96	0.327	0.099	0.107	18.02	0.483	0.094	0.311	15.06	0.156	-0.005	0.204

Notes: Variance and correlation of fixed effects estimated by AKM model as explained in Section IV. Only men are included in these statistics. Only firms and individuals in firms with at least 20 employees are included. Only employed individuals aged 20 to 60 are included in all statistics, where "employed" is defined as earning the equivalent of 2013 minimum wage, adjusted for inflation with the PCE for 40 hours per week in 13 weeks. Individuals and firms in public administration or educational services are not included. Raw decomposition refers to the between- and within-firm variance composition simply on log wages, rather than using the CHK components.

FIGURE A.19 – Correlation of Worker and Firm Effects by Period by Major Industry from 1980 to 2013



Notes: Only men are included in these statistics. Only firms and individuals in firms with at least 20 employees are included. Only employed individuals aged 20 to 60 are included in all statistics, where "employed" is defined as earning the equivalent of 2013 minimum wage, adjusted for inflation with the PCE for 40 hours per week in 13 weeks. Individuals and firms in public administration or educational services are not included. Raw decomposition refers to the between- and within-firm variance composition simply on log wages, rather than using the CHK components. The 7-year periods are 1980-1986, 1987-1993, 1994-2000, 2001-2007, and 2007-2013.

Overall, the findings from the figures corroborate and strengthen our core results from the detailed variance composition in Table IV. 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 are similar to 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.

E Non-Parametric Counterfactual Decomposition: Technical Details

In this appendix section, we describe the technical details underlying the counterfactual decomposition presented in Section III.C. This methodology was developed by Machado and

Mata (2005) and Autor et al. (2005), but is adapted slightly for our purposes.

We start with one observation per person in a given year. For each person, we make note of their log earnings in that year, and their firm's mean log earnings in the same year. Individuals are then sorted into 100 firm-based bins with equal numbers of people, denoted f, on the basis of their firm's mean log earnings. (Thus, except when a firm is right on the border between bins, all people within a given firm are in the same firm-based percentile.) Next, people in each firm-based percentile are sorted into 500 individual-based bins, denoted i on the basis of their own log earnings. There are then 50,000 firm-individual bins, denoted fi, and each person is placed in one of them. So, for example, if $f = 60$ and $i = 400$, that indicates that the bins includes everyone between the 59th percentile and 60th percentile, among all people, in terms of their firm's mean log earnings; and, within that bin, they are between the 79.8th percentile and the 80th percentile by their own earnings.

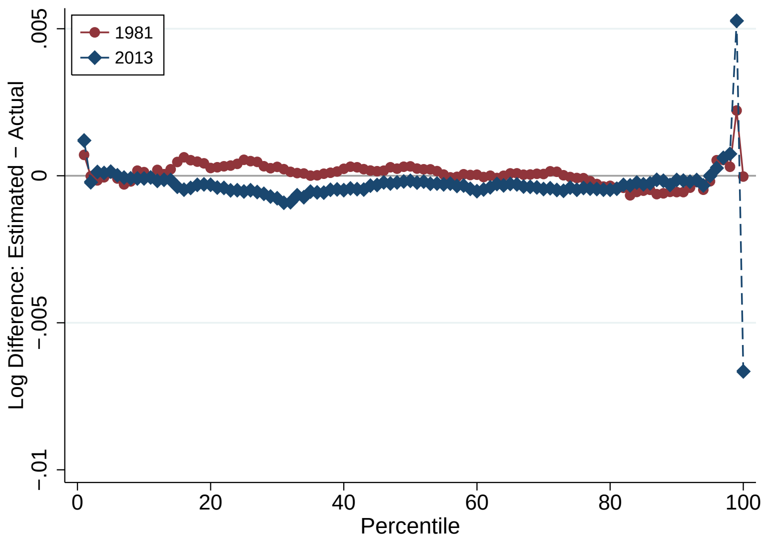
In year t, the mean log earnings in a firm-individual bin is denoted as $\bar{w}_{tfi}$, while the mean log earnings for the entire firm-based bin is denoted by $\bar{w}_{tf}$. Then, a statistic d_{tfi} is calculated for each of the firm-individual bins: deviation from firm bin average, $d_{tfi} = \bar{w}_{tfi} - \bar{w}_{tf}$.

We can now simulate various counterfactuals with the 50,000 bin-level observations. First, we can simulate the actual data for year t with 50,000 data points: each point in this counterfactual is calculated as $c_{fi}^{t,t} = d_{tfi} + \bar{w}_{tf}$ for all fi. (Note that in this case, $c_{fi}^{t,t} = \bar{w}_{tfi}$.) Once we have these 50,000 data points, we then sort them into 100 percentiles, we and calculate the average value of $c_{fi}^{t,t}$ within each percentile, as with the dashed-diamond line in Figure IV. The "change" (in this case just due to the binning procedure) can then be calculated as the difference between the values in these percentiles, minus the average earnings within actual percentiles of the earnings distribution in year t. These changes, calculated based on 1981 data and 2013 data, are shown in Figure A.20. If the binning procedure were perfect, each point would be at zero; in fact, except for some small deviations at the top, they are very close to zero. For example, in 1981, the average log earnings for those in the 99th percentile was 11.9634 (corresponding to \$156,906). Using our counterfactual procedure, the average of $c_{fi}^{1981,1981}$ within the 99th percentile of that statistic was 11.9656 (\$157,251); the difference of 0.0022, the largest such difference for 1981, is plotted in Figure A.20 at the 99th percentile point along the "1981" line.

More interestingly, these statistics allow us to simulate what the distribution would be if between-firm inequality stayed constant at levels from year t, but within-firm inequality changed to the levels observed in year s. To do this, we would use 50,000 data points made up of $c_{fi}^{t,s} = d_{tfi} + \bar{w}_{sf}$ for all fi. Alternatively, we can simulate the distribution if within-firm inequality stayed constant at levels from year t but between-firm inequality changed to the levels observed in year s by using 50,000 data points made up of $c_{fi}^{s,t} = d_{sfi} + \bar{w}_{tf}$ for all fi. In either case, we can compare this to the true distribution in year t or s by sorting these 50,000 data points into percentiles by their value, calculating the average in each percentile, and then comparing the values in these bins to the percentiles of the actual distribution in year t or s.

The results of these counterfactuals are shown in Figure IV. For example, the average of $c_{fi}^{2013,1981}$ within the 99th percentile of that statistic is 12.31; the difference between that value and the average log earnings within the 99th percentile in 1981 (11.96, as noted above) is 0.34; that difference is plotted in Figure IV, on the "Between-Firm Effects Only" line, at the 99th percentile point. Similarly, the average of $c_{fi}^{1981,2013}$ within the 99th percentile of that statistic

FIGURE A.20 – “Counterfactual” difference in distribution due to binning procedure



Notes: Each point shows the difference in average log earnings within that percentile between actual earnings in that year, and earnings simulated in that year using the counterfactual procedure discussed in Appendix E. Only firms and individuals in firms with at least 20 employees are included. Only employed individuals aged 20 to 60 are included in all statistics, where “employed” 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 point shows the difference in average log earnings within that percentile between actual earnings in 1981, and another distribution.

is 12.13; the difference of 0.17 is plotted on the “Within-Firm Effects Only” line, at the 99th percentile point.

References

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

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

Andrews, M. J., L. Gill, T. Schank, and R. Upward, "High wage workers and low wage firms: negative assortative matching or limited mobility bias?," *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 2008, *171* (3), 673–697.

Andrews, M.J., L. Gill, Thorsten Schank, and Richard Upward, "High wage workers match with high wage firms: Clear evidence of the effects of limited mobility bias," *Economics Letters*, 2012, *117* (3), 824–827.

Autor, David H., Lawrence F. Katz, and Melissa S. Kearney, "Rising Wage Inequality: The Role of Composition and Prices," NBER Working Paper No 11628 2005.

Bakija, John, Adam Cole, and Bradley Heim, "Jobs and Income Growth of Top Earners and the Causes of Changing Income Inequality: Evidence from U.S. Tax Return Data," Working Paper, Williams College 2012.

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

Burkhauser, Richard and Kosali Simon, "Measuring the impact of health insurance on levels and trends in inequality," Working Paper 15811, National Bureau of Economic Research 2010.

Card, David, Jörg Heining, and Patrick Kline, "Workplace Heterogeneity and the Rise of West German Wage Inequality," *The Quarterly Journal of Economics*, 2013, *128* (3), 967–1015.

Flood, Sarah, Miriam King, Steven Ruggles, and J. Robert Warren, *Integrated Public Use Microdata Series, Current Population Survey: Version 4.0. [Machine-readable database].* Minneapolis: Minnesota Population Center 2015.

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

Guvenen, Fatih, Greg Kaplan, and Jae Song, "The Glass Ceiling and The Paper Floor: Gender Differences Among Top Earners, 1981–2012," Working Paper, University of Minnesota 2014.

Håkanson, Christina, Erik Lindqvist, and Jonas Vlachos, "Firms and Skills: The Evolution of Worker Sorting," Working Paper 2015:9, IFAU - Institute for Labour Market Policy Evaluation 2015.

Handwerker, Elizabeth Weber, "Increased Concentration of Occupations, Outsourcing, and Growing Wage Inequality in the United States," Working paper, US Bureau of Labor Statistics 2015.

Juhn, Chinhui, Kevin M Murphy, and Brooks Pierce, "Wage Inequality and the Rise in Returns to Skill," *Journal of Political Economy*, June 1993, *101* (3), 410–42.

Kaestner, Robert and Darren Lubotsky, "Health Insurance and Income Inequality," *Journal of Economic Perspectives*, May 2016, *30* (2), 53–78.

Kremer, Michael and Eric Maskin, "Wage Inequality and Segregation by Skill," NBER Working Papers 5718, National Bureau of Economic Research, Inc August 1996.

Lachowska, Marta, Alexandre Mas, and Stephen A. Woodburry, "Sources of Displaced Workers' Long - Term Earnings Losses," 2017.

Machado, José A. F. and José Mata, "Counterfactual Decomposition of Changes in Wage Distributions Using Quantile Regression," *Journal of Applied Econometrics*, 2005, *20*, 445–465.

Piketty, Thomas and Emmanuel Saez, "Income Inequality in the United States: 1913-1998," *Quarterly Journal of Economics*, 2003, *118* (1), 1–39.

Smith, Matthew, Danny Yagan, Owen Zidar, and Eric Zwick, "Capitalists in the Twenty-First Century," Working Papers 2017.

Weil, David, *The Fissured Workplace*, Cambridge, MA: Harvard University Press, 2014.