It's Where You Work: Increases in the Dispersion of Earnings across Establishments and Individuals in the United States

Erling Barth, *Institute for Social Research Oslo and National Bureau of Economic Research*

Alex Bryson, *University College London and National Institute of Economic and Social Research*

James C. Davis, *US Census Bureau, Boston Census Research Data Center, and National Bureau of Economic Research*

Richard Freeman, *Harvard University, National Bureau of Economic Research, and London School of Economics*

This paper analyzes the role of establishments in the upward trend in dispersion of earnings that has become a central topic in economic analysis and policy debate. It decomposes changes in the variance of log earnings among individuals into the part due to changes in earnings among establishments and the part due to changes in earnings within establishments. The main finding is that much of the 1970s–2010s increase in earnings inequality results from increased dispersion of the earnings among the establishments where individuals work. Our results direct attention to the role of establishment-level pay setting and economic adjustments in earnings inequality.

We have benefited from support from the Labor and Worklife Program at Harvard University, the National Bureau of Economic Research, and the Norwegian

[*Journal of Labor Economics*, 2016, vol. 34, no. 2, pt. 2]
 0734-306X/2016/34S2-0012$10.00
Submitted July 11, 2014; Accepted May 18, 2015; Electronically published February 11, 2016

The defining feature of the distribution of US earnings from the mid-1970s through the 2000s is the huge increase in inequality. Analysis of individual earnings shows that inequality increased among workers with different observed measures of skill, such as education, age, and occupation, and that earnings increased more at higher percentiles of the earnings distribution than at lower percentiles, even among workers with the same measured skill (see, e.g., Lemieux 2006; Autor, Katz, and Kearney 2008).

This paper examines earnings inequality along a dimension that previous research has largely ignored: the establishments that employ the worker. Viewing inequality through an establishment lens, we find that most of the increased variance in earnings among individuals is associated with the increased variance of average earnings among the establishments where they work. Our findings direct attention to the role of establishment and firm pay-setting and labor market adjustments by place of work in the rise in inequality.[1]

To analyze the effect of establishment earnings on the trend increase in inequality, we combine several data sets: the March Current Population Surveys (CPS) files that record annual earnings and weeks worked of individual workers; the US Census Bureau's Longitudinal Business Database (LBD), which is the longitudinal version of the Census Business Register with data on establishment payroll and employment (see Jarmin and Miranda 2002); and the Longitudinal Employer-Household Dynamics database (LEHD), which contains data on the earnings of millions of workers and their places of work from Unemployment Insurance files. We link the LBD and the LEHD through establishment identifiers and use the linked data file to decompose the inequality of earnings among workers into the part that occurs among establishments and the part that occurs within establishments. Since the LEHD data do not include information on individuals'

Research Council (project 173591/S20; Barth and Bryson). Opinions and conclusions expressed herein are those of the authors and do not necessarily represent the views of the US Census Bureau. Results have been reviewed to ensure that no confidential information is disclosed. Contact the corresponding author, Richard Freeman, at freeman@nber.org. Information concerning access to the data used in this article is available as supplementary material online.

[1] Previous work on the employer's role in wage setting includes Davis and Haltiwanger (1991), Groschen (1991), Abowd, Kramarz, and Margolis (1999), Hellerstein, Neumark, and Troske (1999), Lane, Salmon, and Spletzer (2007), and Gruetter and LaLive (2009), following the early works on interindustry wage differentials, Dickens and Katz (1987), Krueger and Summers (1988), Bell and Freeman (1991), and Gibbons and Katz (1992). Lazear and Shaw (2009) made the observation that across-firm differences appeared to be growing over time for a significant number of countries, as, for example, seen in the contribution on Sweden by Nordström Skans, Edin, and Holmlund (2009) in their volume. Card et al. (2013) find a growing contribution of plant heterogeneity in wages in Germany between 1985 and 2009.

education, we link individuals on the LEHD to their responses on the 1990 and 2000 Census long-form sample and 1986–98 March CPS files to determine workers' years of schooling.

Section I of the paper estimates the contribution of changes in the dispersion of average earnings across establishments to the rise in inequality. Section II connects the distribution of establishment earnings to returns to measured skills and individual characteristics and to the sorting of workers among establishments. Section III estimates the contribution of establishment earnings to the growth of earnings at each percentile of the earnings distribution and to the increased gap between top earners and other workers. Sections IV and V assess the pathways behind the widening distribution of establishment level earnings. Section VI concludes.

I. Earnings among Establishments and Earnings Inequality among Workers

Analysis of the link between growing earnings inequality among workers and changes in the distribution of earnings among establishments requires earnings data for individuals and establishments and links between individual and establishment earnings. We measure individual earnings by weekly earnings (annual earnings/weeks worked) from the internal US Census version of the March CPS files,[2] and we use the variance of log weekly earnings as our measure of inequality. The internal Census CPS has higher top codes for income and thus more accurate earnings at the top of the distribution than publicly available files.[3] We measure establishment earnings by annual earnings per worker (payroll before deductions/number of employees) in the LBD, and we measure inequality by the variance of log annual earnings per worker.

Panel A of figure 1 displays estimates of the variance of log weekly earnings for individuals from the March CPS and the variance of log annual earnings among establishments from the LBD. The top line shows a substantial increase in the variance of March CPS earnings that is similar in magnitude to increases found in other CPS-based earnings data. The middle line gives the variance of log average earnings among establishments, weighted by establishment employment to be comparable with the CPS variance for individuals. The variance of establishment earnings lies below the variance of individual earnings because the establishment variance excludes variation within establishments, whereas the variance of individual earnings includes the variance within establishments as well as among

[2] The pattern of change in log weekly earnings resembles the pattern in the widely studied log hourly earnings from the CPS Outgoing Rotation group files. Lemieux (2006) compares CPS-based inequality measures.

[3] See the data section in the appendix. The top coding in the internal CPS affects less than half a percent of the sample. See Burkhauser et al. (2011). The LBD and LEHD data are not top-coded.

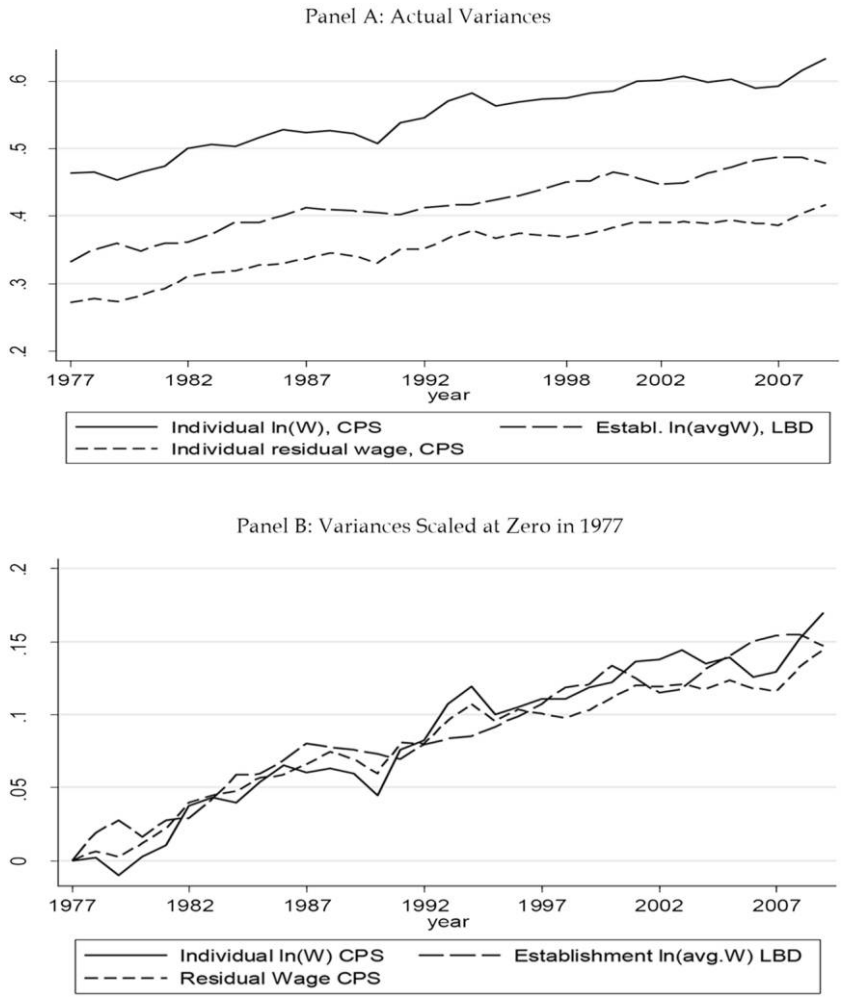

FIG. 1.—Variance of log earnings: individuals and establishments, 1977–2009: *A*, actual variances; *B*, variances scaled at zero in 1977. The upper panel shows the level of the variance of log(earnings) from 1977 to 2009, while the lower panel shows the accumulated change in the variance for the same period, with 1977 set to zero. The variance of log(weekly earnings) is calculated over individuals from the March Current Population Surveys (CPS) and over establishments' average wages from the Longitudinal Business Register Data (LBD) (employment weighted). The CPS residual wage is calculated from yearly regressions of individual log (weekly earnings) on years of education, experience (Mincer 1974), experience squared, and a race dummy, all interacted with gender. See the appendix for details and table A2 for CPS results for the LEHD states, weekly versus hourly earnings, and Gini coefficients, and for measures of relative wages in tens of decile ratios. The LBD data are detailed further below.

establishments. The bottom line gives the residual variance from regression estimates of log earnings on the worker characteristics specified in the table note. Reflecting the role of human capital and demographic factors in earnings, the residual variance lies below the unadjusted variance among individuals and below the variance of establishment earnings as well.

To focus attention on the similarity in changes among the three measures, panel B of figure 1 displays the accumulated change in variances of log earnings over time, scaled at 0 in 1977. The 1977–2009 increase for individual earnings is 0.170 log points. The increase in the earnings equation residuals is 0.147 log points. These estimates imply that 86% (0.147 points/0.170 points) of the overall trend is due to the residuals, while 14% is associated with the observable attributes of workers.[4] The variance of establishment earnings increased by the same 0.147 points as the variance of residuals. Thus, if we take the increased variance in establishment earnings and the 0.023 point increase in the variance due to observable worker attributes, we get the entire increase in the variance of individual earnings. The exact accounting is happenstance, but the calculation demonstrates our main finding: that increased variance of establishment earnings is a major pathway for the increased variance in individual earnings.[5]

Given that the variances in figure 1 come from different earnings series, the analysis falls short of an ANOVA decomposition of the trend increase in inequality into its between-establishment and within-establishment components. An ANOVA requires a single earnings series with identifiers for individuals and establishments, which the LBD and CPS do not have. The absence of data on the earnings of workers in establishments manifests itself in our estimate of the variance of establishment earnings. Absent earnings with individual and establishment identifiers, we use the variance of the *log average* establishment earnings instead of the variance of the *average log* worker earnings in an establishment appropriate to a complete variance decomposition.

How much does this distort the calculations? To estimate the magnitude of the distortion, we applied Aitchison and Brown's (1957, 8) formula for the difference between the variance of log average establishment earnings and the variance of the average of log earnings when data are distributed log-normally.[6] Appendix table A1 estimates the differences in the

[4] Age and education explain most of the 14% of the increased variance due to observable worker attributes. See Sec. II.

[5] See Davis and Haltiwanger (1991) and Dunne et al. (2004) for early observations of this relationship for manufacturing establishments.

[6] In the appendix, we use LEHD data to adjust the variance of log average establishment earnings to approximate the variance of the log of average earnings using $\ln E(w) = \mu_f + \sigma_f^2/2$, where μ_f is $E[\ln(w)]$ and σ_f^2 is the within-establishment variance of log earnings. The 1992–2007 variance increase is 0.070 (adjusted) and 0.075 (unadjusted).

two variances and finds only modest differences in the levels of the variances and virtually identical changes in the variances over time. As long as the log-normal assumption holds, using the variance of log average earnings rather than the variance of the average log of earnings for the establishment variance does not substantively distort the figure 1 results. To examine this further without the log-normal assumption, we computed inequality measures over time from the CPS for individuals and from the LBD for establishments and also found that the increased dispersion of earnings among establishments is close in magnitude to the increased dispersion of earnings among individuals. Table A2 reports these results for decile measures of dispersion and for the Gini coefficient calculated on CPS individual earnings and LBD establishment average earnings. The patterns of change are again similar between the establishment averages and individual earnings.

A. LEHD Earnings

The LEHD allows us to do better by linking earnings to individual workers and to the establishments where they work,[7] which is necessary for a genuine ANOVA decomposition of log earnings into its between- and within-establishment components. In this analysis, we measure individual earnings by yearly earnings for workers employed in all four quarters of a year in the nine states that provide such information from 1992 to 2007.[8]

To see if the LEHD earnings are representative of the United States, we compared the variance of log yearly LEHD earnings to the variance of log March CPS weekly earnings for the nine states. We obtained similar levels of variance and nearly identical changes in variances.[9] We then compared the CPS variance of log earnings in the nine states to the variance for the whole country and also found similar levels and nearly identical changes.[10] Thus,

[7] The LEHD and LBD link identifies the firm that employs workers and the establishment in which they work when firms have one establishment in a locality. When firms have multiple establishments in a locality, the Census uses a probabilistic worker assignment to estimate the establishment in which the worker was employed. We use the Census's probabilistic assignment to identify the establishment location of all workers. See Abowd et al. (2002, 2003, 2007) for details and methods regarding the use of LEHD data.

[8] See the the data section of the appendix for details.

[9] The LEHD variance for 1992 is 0.506; it is 0.588 for 2007 (table 1). The CPS-based variance for the same states is 0.538 in 1992 and 0.618 in 2007. The increases in the LEHD-based variance (0.082) and CPS-based variance (0.080) are also nearly identical.

[10] Appendix table A2 gives a CPS-based variance of log earnings for the United States of 0.546 in 1992 and 0.633 in 2009. The CPS-based variance for our nine states is 0.538 in 1992 and 0.623 in 2009. The 1992–2007 change for the United States (0.087) is almost identical to that for the nine LEHD states (0.085).

analysis of the LEHD for the nine states should generalize to the entire country.[11]

Given these assurances, we decomposed the LEHD earnings into their within- and between-establishment components and calculated changes in the components over time. Denote $\ln w_{ip}$ as the log earnings of worker i in establishment p, $E\ln w_{ip}$ as the mean log *earnings for workers in establishment p*, V_w as the within component of variance, and, V_b as the between component. The variance decomposition of log earnings is

$$V(\ln w_{ip}) = V_w + V_b = V(\ln w_{ip} - E\ln w_{ip}) + V(E\ln w_{ip}). \quad (1)$$

Table 1 records the decomposition from 1992 to 2007. The between-establishment earnings are weighted by their employment so that they give greater weight to establishments with more workers. In both 1992 and 2007, log earnings varied more within establishments than among establishments. But the increase in the between-establishment variance (0.056) is over twice the increase in the within-establishment variance (0.027), so that the between component accounts for 67.5% of the increased variance among all workers.[12] The 67.5% estimate falls short of figure 1's 87% estimated establishment share, but it is sufficiently large to support the claim that increased inequality among establishments is the major pathway for increased inequality among workers.

B. Stayers

The longitudinal nature of the LEHD allows us to estimate the relation between the dispersion in average establishment earnings and dispersion in individual earnings in another way. This is by decomposing the change in the variance of log earnings for workers who *stay* in the same establishment from one year to the next into its between- and within-establishment components. Analysis of changes in inequality among *stayers* holds fixed the time-invariant characteristics of workers and establishments, including characteristics for which we have no observable measures. This analysis pins down the impact of the widening establishment earnings on individual earnings in a way that sidesteps complications due to the connections between earnings, labor mobility, exit and entry of establishments, and matching of workers and establishments.

[11] We also examined the pattern of change in other states that the LEHD covered over shorter periods and found similar results to those in our sample of states.

[12] The calculation is 0.056 points/0.083 points = 67%. The results are similar if we take earnings for the larger sample of workers who appear in at least a single quarter (the second quarter of the year in our calculation). They are also similar for 22 states that appear in the data for a shorter time period. Results are available from the authors on request.

Table 1
Level and Changes in Variance in Log Earnings Between and Within Establishments, Nine LEHD States, 1992–2007

	1992	2007	Growth
Variance across individuals, total	.480	.563	.083
Between establishments	.219	.275	.056
Within establishments	.260	.287	.027
No. of individuals (millions)	19.0	26.0	
No. of establishments (millions)	1.33	1.81	

NOTE.—The table shows annual earnings for full-year employees in the main job. Results for quarterly earnings for all individuals observed at the employer in the second quarter, as well as figures including 22 states for shorter periods of time, show similar patterns and are available from the authors on request.

To see how data on stayers illuminates the role of establishments, consider two establishments, all of whose workers are stayers. In this case, inequality of worker earnings could increase because of (i) increased earnings differentials between the establishments, with unchanged relative earnings within establishments, (ii) increased relative earnings within establishments, with unchanged differentials between establishments, or (iii) some mixture of between- and within-establishment changes. The decomposition for stayers arithmetically measures the between-establishment and within-establishment effects on stayers' inequality.

Row 1 of table 2 gives our estimates of the *change in variance* of log earnings for stayers from year $t - 1$ to t over specified periods. Since workers who stay with an establishment differ from one year to the next, we maximize the number of persons in the computation by using a rolling sample. We calculated log earnings for stayers in years $t - 1$ and t, computed the variance in both years, and then took the change in variances from $t - 1$ to t to measure the change in variance. We repeated the calculation for year

Table 2
Growth in Variance Components Within and Between Establishments: Stayers and All Employees, LEHD Data, 1992–2007

	Period of Change		
	1992–97	1997–2002	2002–7
Stayers:			
Change in Var(log earnings)	.013	.011	.037
Change in between variance	.013	.008	.027
Change in within variance	.001	.003	.009
All employees:			
Var(log earnings)	.023	.020	.040
Between	.015	.012	.029
Within	.007	.009	.011

NOTE.—The table shows the accumulated growth in the variance of log(earnings) each 5 years from 1992. Employment-weighted means for the nine LEHD states; see the the data section of the appendix. The top panel shows the accumulated change calculated on year-to-year stayers only; the bottom panel shows growth for all.

t to $t + 1$ and ensuing years. The 0.013 in the column labeled 1992–97 sums the change in the variance of log earnings for stayers from 1992–93 to 1996–97. The 0.024 in the 1997–2002 column sums the change in variance from 1997–98 to 2001–2, and so forth. The estimates show moderate increases in variance in 1992–97 and 1997–02 followed by a larger increase in 2002–7. Over the full period, the change in variance is 0.061 log earnings points.

How much of the changed variance among stayers is associated with changes in earnings among establishments? The row "Change in between-establishment variance" estimates the change in variance of the average log earnings among establishments. These estimates are the sum of the changed variance of establishment-level log earnings of stayers from one year to the next over the specified period. They attribute all of the increased variance among stayers from 1992 to 1997 to the increased between-establishment variance (0.013 points/0.013 points) and attribute smaller but still dominant shares of the increased variance in ensuing periods to the increase in variance among establishments. For the whole period, the 0.048 change in variance due to the changed variance among establishments is 79% of the 0.061 total increase in variance. The remaining 21% of the increase in total variance is the contribution of changes in within-establishment variance.

The bottom part of table 2 summarizes the results of an analogous variance decomposition for all employees. Changes in variance are larger for all employees than for stayers because all employees are a more heterogeneous group that includes workers who move from one establishment to another or between employment and nonemployment. The variance among all workers increases by 0.083 points, of which two-thirds (0.056/0.083) is between establishments. Dividing the change in total variance for stayers by the change for all employees shows that the stayers account for nearly three-quarters of the increased overall variance. This reflects the fact that most workers stay in the same job from one year to the next. Exit and entry of establishments and movement of workers among establishments and between work and nonwork contribute to the variance, but the main driver of the trend in variance for all workers is the increased variance among stayers due to changing establishment differentials.

II. Worker Characteristics and Establishment Premium

Most studies of earnings inequality focus on the contribution of increased returns to observable characteristics, such as education or age. To examine the interaction between establishment earnings and the returns to individual characteristics and sorting of workers by these characteristics among establishments in the rising trend in inequality requires a valid measure of years of schooling, which the LEHD does not provide. To obtain a measure of schooling for individuals, we matched the LEHD records to the 1990 and 2000 US Census long forms and the 1986–98 March CPS files to

obtain Census or CPS years of schooling to add to the LEHD.[13] We then estimated the following extension of the standard log earnings equation each year from 1992 to 2007:

$$\ln w_{ip} = x_{ip}b + \varphi_p(i) + u_{ip}, \text{ with } E\left(u_{ip} \mid x_{ip}, \varphi_p\right) = 0. \quad (2)$$

In this equation x_{ip} is a vector of worker characteristics (years of schooling, experience and its square (Mincer), and dummy variables for nonwhite and gender) for worker i in establishment p. We interact the independent variables with gender to allow for male-female differences in the relation of the independent variables to log earnings.

Our extension of the standard log earnings model is the vector of dummy variables $\varphi_p(i)$ for the establishment p where the worker i works. We impose equal establishment effects on workers by omitting the individual subscript from establishment dummy variables, but we still write the vector as a function of i to highlight that all workers in an establishment share the same establishment effect. This specification puts individual heterogeneity in the establishment effect (which reflects the quality of the individual and establishment match) into the error term.

Taking the variance of (2), we decompose the variance of log earnings into the part due to variance of the predicted wage from observable characteristics among workers, the variance of earnings among establishments, the covariance between them, and the variance in the error term. To simplify the algebra, denote a worker's predicted wage from observable characteristics as s ($= xb$, a composite index that depends on worker attributes weighted by the estimated b coefficients linking attributes to earnings), and denote $V(\varphi)$ as the variance of the establishment's effect on wages. This yields

$$V(\ln w) = V(s) + V(\varphi) + 2\text{Cov}(s, \varphi) + V(u). \quad (3)$$

Denoting S as the establishment's average level of the predicted wage from observable characteristic(s), a natural measure of the similarity of workers in an establishment is $\rho = \text{Cov}(s, S)/V(s)$. The ρ coefficient is Kremer and Maskin's (1996) index of worker-worker segregation across establishments. When establishments hire workers randomly by observable characteristics, $\rho = 0$. When workers are perfectly sorted with workers having similar characteristics, $\rho = 1$. Similarly, we measure the extent to which the attributes of individuals that contribute to wages are associated with the establishment effect by $\rho_\varphi = \text{Cov}(s, \varphi)/V(s)$. When firms hire workers by observable characteristics independently of their establishment earnings premium, $\rho_\varphi = 0$.

[13] The long form is distributed to approximately 15% of the US population every decennial. The combination of the Census long form and the CPS allows us to match 18% of the LEHD sample with those files and thus obtain valid education measures for a large number of workers. See the appendix for details.

Given these definitions, the between-establishment variance divides into a part due to sorting of workers and a part due to "pure" variation of earnings among establishments:

$$V^b = V(s)\left(\rho + 2\rho_\varphi\right) + V(\varphi), \tag{4}$$

where $V(s)(\rho + 2\rho_\varphi)$ reflects the contribution to between-establishment variance of both forms of sorting of workers and where $V(\varphi)$ is the variance of the establishment effect for workers with similar measured characteristics independent of variation in the distribution of workers among establishments.

Similarly, we decompose the within-establishment part of the variance V^w into

$$V^w = V(s)(1 - \rho) + V(u). \tag{5}$$

When establishments employ workers with the same characteristics, $\rho = 1$ and the variance of the individual wage index contributes nothing to within-establishment variance. When establishments hire workers irrespective of characteristics, $\rho = 0$. In this case, the variance in the distribution of individual attributes contributes to the within-establishment variance but not to the across-establishment variance.

Table 3 gives our decomposition of earnings in the matched LEHD-Census sample. The earnings equation (2) is estimated separately for each of the nine LEHD states, and the table reports employment-weighted statistics across the states. The "Var $(\ln w)$" row records the variance of log earnings. The variances for the matched sample are similar to the table 1 variances for the entire LEHD, with a slightly larger increase.[14] The similarity shows that the matching preserved the pattern of change in dispersion on which we focus.

The row "Individual characteristic: Var(s)" shows that the variance of the predicted individual earnings, conditional on establishment effects, had a negligible effect on the trend in variance. Since the education premium was widening (Goldin and Katz 2008), something else in the index of worker characteristics must have offset its effect on the variance. As we shall see, that something else is a fall in male/female earnings differences.

The estimated sorting coefficients show that worker-worker sorting (ρ) increased by 1.3 percentage points over the 15-year period. Worker-establishment sorting, ρ_φ, increased by a larger 6.5 percentage points, as establishments with high earnings increasingly loaded up on high-pay workers. But because the sorting effect depends on the variance of the index of individual characteristics, $V(s)$, which fell slightly, sorting has little impact in the decomposition.

[14] There is an increase of 0.088 in table 3 as compared to 0.083 in table 1.

Table 3
Variance Decomposition of LEHD Earnings with Individual Characteristics

	1992	1997	2002	2007	1992–2007 Change	Share of Change
All:						
Var ($\ln w$)	.457	.478	.500	.545	.088	1.00
Individual characteristic: Var(s)	.108	.101	.101	.101	−.007	−.08
Worker-worker: ρ	.344	.340	.345	.357	.013	
Worker-establishment: ρ_{φ}	.233	.242	.258	.297	.065	
Variance between	.235	.246	.259	.292	.057	.65
Establishment effect: $V(\varphi)$	.147	.162	.172	.196	.049	.56
Individual characteristic contribution: $V(s)\rho$	.037	.035	.035	.036	−.001	−.01
Match contribution: $V(s)2\rho_{\varphi}$	.050	.049	.052	.060	.010	.11
Variance within	.223	.232	.241	.253	.031	.35
Within residual: $V(u)$	.152	.165	.174	.189	.037	.42
Individual characteristic contribution: $V(s)(1 - \rho)$	.071	.067	.066	.065	−.006	−.07
No. of individuals (millions)	3.9	4.2	4.3	4.3		
No. of establishments (millions)	.7	.8	.8	.8		

NOTE.—Regressions are estimated on the matched Census LEDH sample, including Decennial 1990 and 2000 and the Current Population Survey sample from the nine LEHD states. For the employment-weighted means for the nine LEHD states, see the appendix. The index of individual characteristics ($s = X\beta$) includes experience (Mincer 1974), experience squared, years of education, and a nonwhite dummy, interacted with gender. Employer identification is employer-state-id-unit (SEIN UNIT). Earnings is obtained from the LEHD data, annual earnings of full-year employees in the main job, while education, age, and race are obtained from the Census long form and the CPS.

What dominates the increased variance of establishment-level earnings is the increased divergence of earnings among establishments. This contributes 0.057 points, or 65%, of the increased variance. In turn, the decomposition of the between-establishment effect shows that the increased variance in the establishment effect, φ_p, accounts for the vast bulk (.049/.057 = 86%) of the increase in the between-establishment variance.

Finally, the decomposition of the within-establishment variance at the bottom of the table shows that the within-establishment increase resulted largely from increased variance of the residual in the equation—that is, to greater variance among workers with similar characteristics within establishments rather than to changes in the within-establishment composition of workers.[15]

[15] Workers in establishments with multiple employers within the same state are allocated to establishments using distance to work vs. residence, together with summing conditions. See Abowd et al. (2002) for the allocation algorithm. To check that our results are not sensitive to this allocation, we report results using SEIN rather than SEIN UNIT as the organizational unit in table A3 in the appendix. The between-employer variance goes down to 0.276 vs. 0.292, and the share of the increase in the variance due to between employer variation is 61% rather than 65% when using state-employer instead of establishment. Otherwise the patterns are very similar.

The surprisingly small (and negative) effect of the variance of the index of individual characteristics on the change in dispersion of earnings within and between establishments merits attention in light of large increases in the estimated coefficient on education, which add to the variance of earnings. To understand what lies behind the small estimated composite effect, we decomposed the variance of March CPS earnings yearly from 1977 to 2011 and calculated the contribution of worker attributes to the overall increase in variance.

Figure 2 gives the results of this decomposition. The line for years of schooling shows that schooling increased the variance of log earnings as the return on years of schooling increased. But the line for gender shows a large decline in the variance of log earnings associated with gender.[16] From 1977 to 2011, the schooling measure added 0.07 points to the variance, while the gender measure reduced the variance by 0.06 points. Over the 1992–2007 period, the more modest upward trend in variance due to schooling is partially offset by declines in variance due to gender, age, and the covariances as well.

III. The Widening Percentile Distribution of Earnings

Studies that focus on the entire distribution of earnings document that percentage changes in earnings in the period on which we focus were larger in the higher percentiles of the distribution and were especially large for top earners—the upper 10% or 1%, depending on the study (Lemieux 2008; Alvaredo et al. 2013).

To see how establishment differentials affect changes in earnings by percentile in the earnings distribution, we calculated LEHD percentile earnings distributions for individuals in 1992 and 2007. We assigned to each person the establishment effects of their workplace and calculated the mean of establishment effect for all individuals at a given percentile.[17] If the distribution of earnings in 1992 had 1,000 workers at the 10th percentile, the establishment effect for the 10th percentile would be the average of the establishment effects for the 1,000 workers. Similarly, if the distribution of earnings in 2007 had 1,500 workers at the 10th percentile (due to the increased work force), the establishment effect for the 10th percentile would be the average of the establishment effects for those workers. Given these estimates, we then calculated the increase in the average establishment

[16] In this calculation, we included the covariance of gender with age. We made similar calculations for the matched LEHD data and obtained similar results. In that data set, adding establishment effects reduces the estimated educational wage differentials by about 20%, reflecting a positive sorting of high-educated workers toward high-paying establishments.

[17] The regression includes years of schooling, experience and experience squared, and a race dummy, all interacted with gender, in addition to an establishment fixed effect.

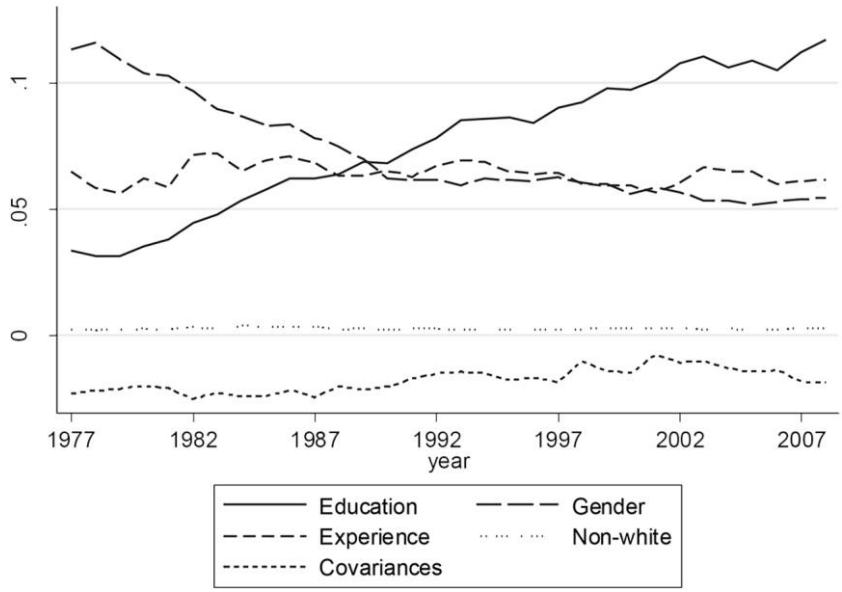

FIG. 2.—Variance decomposition of log earnings from CPS, 1977–2009, based on estimated impacts of individual characteristics from yearly regressions. Calculated from yearly regressions of individual log(weekly earnings) on years of education, experience, experience squared, and a nonwhite dummy, all interacted with gender. Each component consists of the sum of the gender-specific terms. The "Gender" line includes the gender dummy and the covariance between age and gender, and the line labeled "Covariances" summarizes the remaining covariance terms.

effects by percentile between 1992 and 2007. If changes in establishment earnings were important in altering the distribution of individual earnings, the pattern of change in the establishment earnings by percentile should closely resemble the pattern changes in the earnings of workers in the percentiles.

Figure 3 shows that this is the case. The dotted line for the changes in the average establishment effect for workers by percentile increases with the percentiles of the distribution. To see how this meshes with the changes in earnings of individuals at each percentile, we calculated the average log earnings of individuals by percentile in 1992 and 2007 and computed the difference between these percentile averages. To better contrast changes for individuals and changes for establishment effects, we then subtracted the average change in log earnings for all individuals from each percentile's change. We did this because by construction the changes in establishment differentials average to zero with negative as well as positive effects. Subtracting the change in the mean for individual earnings preserves relative changes while putting individual changes in similar units as the establishment changes.

FIG. 3.—Change in average log earnings, by percentile of the earnings distribution among individuals, 1992–2007. The horizontal axis is the percentile of the distribution of individual earnings. The vertical axis shows the difference between average log earnings from 1992 to 2007 for each percentile. The solid line shows changes in individual earnings, while the dashed line shows the change in average establishment effects of the individuals in each percentile. The establishment effect is the estimated establishment fixed effect from yearly log earnings regressions on education, experience (Mincer 1974), experience square, and a race indicator, all interacted with gender, and establishment dummy variables. Data from LEHD using the Census 1990–2000 and CPS sample of nine LEHD-states, annual earnings, full-year employees, main job.

The solid line in figure 3 shows these changes. The pattern of changes for individual earnings and for establishment effects closely mirror each other. Establishment effects have larger increases than individual earnings at the lower end of the distribution and smaller increases than individual earnings at the top percentiles. These differences reflect the fact that the earnings distribution is ordered by individuals, whose changes will be influenced by their circumstances as well as by establishment effects—individuals low in the distribution will have negative shocks and those high in the distribution will have positive shocks. But the deviations are modest. Changes in earnings at the establishment where people work dominate the pattern of higher increases in earnings at higher percentiles of the distribution.

Finally, given widespread attention to the increased relative rewards to workers at the top of the earnings distribution, we examined the extent to

which the advantage at the top increased because earnings at the establishments at which they work increased relative to earnings at other establishments. We divided the LEHD sample into top earners—defined as those in the upper 5% of the distribution of the nine states—and the remaining 95%. We computed the 1992–2007 increase in the log earnings difference between the top 5% and the 95% and the impact that increase had on earnings inequality for all workers. We then estimated the change in earnings at the establishments where the top 5% worked relative to the establishments where the 95% worked and the impact that had on the difference between the 5% and the 95%.

Table 4 shows that the increased advantage of the 5% accounts for 40% of the increase in the variance of log earnings measures of inequality and that the divergence of establishment earnings underlies much of the increased advantage of top earners. Row 1 records the variance of log earnings and change in variance for all workers in 1992 and 2007. Rows 2 and 3 estimate the log mean earnings and changes in log means for the top 5% and the remaining 95%. Row 4 gives the differences in the means. The earnings advantage of the top 5% over the 95% increased by 0.208 log earnings points. Row 5 uses the variance formula in the table note to calculate the impact of the earnings gap to the total variance in each year and the impact of the increase in the gap to the increased variance for all workers. It is

Table 4
Effect of Increase in Top 5% Earners/Other Earners Gap to Inequality and of Increased Establishment Differentials on Top 5%/Other Earners Gap

	1992	2007	Change
Contribution of earnings gap between upper to variance:			
1. Variance of log earnings, all workers	.480	.563	.083
2. Mean, log earnings, upper 5%	7.843	8.142	.299
3. Mean, log earnings lower 95%	6.261	6.352	.091
4. Difference in means (row 2 − row 3)	1.582	1.790	.208
5. Contribution of difference in means to variance	.119	.152	.033 (40% of row 1)
Impact of establishment effects:			
6. Establishment effects, 95th percentile	.465	.630	.165
7. Establishment effects, below 95th percentile	−.024	−.033	−.009
8. Difference in establishment effects (row 6 − row 7)	.489	.663	.174 (84% of row 4)

NOTE.—Data are from the nine LEHD states, 1992–2007 (employment-weighted means). The contribution of the difference in means follows arithmetically from decomposing the variances of log earnings into differences in the means between the two groups and the variances within the groups. If $E(5\%)$ is the mean log earnings of the top 5%, and $E(95\%)$ is the mean log earnings of the remaining 95%, and $V(5\%)$ is the variance of log earnings within the top 5%, and $V(95\%)$ is the variance of log earnings within the remaining 95%, the variance of log earnings for all workers V decomposes into $(.95)(.05)(E5\% - E95\%)^2 + .95V(95\%) + .05(V(5\%)$.

from this calculation that we obtain the 40% figure cited above as the effect of the changed gap on the total increase in variance between 1992 and 2007.[18]

The remainder of the table assesses the role of changes in establishment differentials on the 0.208 increased advantage of the top 5%. Rows 6 and 7 estimate the establishment effects for the 5% and for the 95%. The estimates follow the procedure in the figure 3 calculations just described: they average the establishment effects from the LEHD earnings regression for all persons in the relevant groups.[19] Note that per the figure 3 discussion, the establishment effects are scaled around zero, which places them on a different metric than the mean earnings in rows 2–4. But the changes over time are comparable. Row 8 shows that the change in the establishment effects for the top 5% versus the 95% was 0.174. This is 84% of the change between the mean earnings of two groups in row 4. Given that 40% of the increased variance of log earnings is associated with the pulling away of the top 5% versus others, the implication is that 33% ($0.84 \times 40\%$) of the increased variance of log earnings is attributable to increased gap between the average earnings in the establishments where the top 5% work and the average earnings in the establishments where others work.

In sum, changes in the distribution of earnings among establishments substantially affected changes in earnings along the entire earnings distribution and the increased advantage of top earners compared to other workers. The question that naturally arises next is, what forces have moved establishments further apart from each other in earnings space?

IV. Pathways for the Widening Earnings Structure among Establishments

To answer this question and to determine what establishment-level factors are associated with the establishment average earnings, we regressed log average yearly earnings in establishments on measures of establishment attributes using the following equation:

$$\ln w_p = G_p a + I_p b + c \ln E_p + d \mathrm{MU}_p + e \ln \mathrm{Ef}_p + f \ln \mathrm{NP} + \varphi_p, \quad (6)$$

where w_p is the average annual earnings in an establishment in year t from the LBD. The vector φ_p measures establishment mean earnings net of the

[18] The 60% of the rest of the increase in variance is due largely to increased variance in log earnings among the 95% is associated with the widening of establishment effects in their establishments.

[19] They come from the same regression of log earnings of individuals on years of schooling, experience and experience squared, a race dummy, interacted with gender and the key vector of establishment dummies that yields the establishment effect.

other variables in the regression. It differs from the establishment effects examined in Sections I–III mainly because it does not contain controls for individual characteristics, such as education, as the LBD has no data on individuals. Variable *G* is a vector of 537 dummy variables for the geographic area in which an establishment locates: for urban areas, it is the metropolitan area (PMSA), and for rural establishments outside of PMSAs, it is the Bureau of Economic Analysis (BEA) economic area. Variable *I* is a vector for the industry in which the establishment's production fits according to the North American Industry Classification System (NAICS) code, which we vary from the one-digit (nine groups groups) to the four-digit level (277 groups). The next set of variables reflect the size of the employing business: variable *E* is the number of employees in the establishment, MU is a dummy for whether the establishment is part of a larger firm that has multiple units/establishments; for those that are multi-unit, Ef is employment in the firm and NP is the number of establishments in the firm.

Table 5 summarizes the results from estimating equation (6). Each row represents a model in which we include industry dummy variables from one digit to four digits, with the final row adding employment size variables as well. The 2007 calculations show that neither geography nor size of the employing business contributes much to the variance in that year. What matters is industry: its contribution rises from 20% to 49% when we expand the number of dummy variables going from one-digit to two-digit

Table 5
Variance and Growth in Variance Decompositions: Establishment-Level Earnings

	Dependent Variable: Log Earning (Establishment Wage)				
	Geographic Unit	Industry	Establishment	2Cov (*I*, *G*)	Employment
Level 2007 (SIC):					
1-digit industry	.05	.20	.75	.00	
2-digit industry	.04	.49	.52	.01	
3-digit industry	.04	.49	.46	.01	
4-digit industry	.04	.52	.43	.01	
4-digit + employment	.03	.52	.42	.01	−.01
Change 1977–2007:					
1-digit industry (SIC)	.04	.23	.72	.00	
2-digit industry	.03	.49	.43	.01	
3-digit industry	.03	.49	.44	.01	
4-digit industry	.03	.52	.41	.01	
4-digit + employment	.03	.52	.40	.04	.00

NOTE.—For different industry detail, the table shows the share of variance (change in variance) attributed to the various factors, based on regression analysis of log(establishment average wage). Geography is defined as PMSA, and outside of the PSMA's, BLS working area within state is used. The number of geographic units is 537. The number of digits refers to the SIC classification (after 1998, industries are classified according to NAICS, 6, 4, 3, 2, 1 digits). Employment includes establishment size, firm size, the number of establishments of the firm, and a dummy for multi-unit firm. The establishment factor is the residual from each regression and is thus not allowed to covary with the other factors.

industries, then increases modestly with additional industrial detail. Establishment effects also matter: they represent 42% of the variance with detailed industry codes and employment and covariances.[20] In 2007, the four-digit industries with the highest pay were Securities and Commodity Contracts Intermediation and Brokerage, Securities, and Commodity Exchanges, Software Publishers, Pipeline Transportation of Crude Oil and Natural Gas, Other Financial Investment Activities, and Electric Power Generation, Transmission, and Distribution. The four-digit industries with lowest pay included Apparel Manufacturing, Food and Beverage Stores, Food Services and Drinking Places, Accommodation, and Book, Periodical, and Music Stores.

The decomposition of the change in variance from 1992 to 2007 shows that industry and establishment also dominate changes over time. Two-digit industry dummies provide considerable information about changes in establishment earnings, but there remains considerable variance in the changes among establishments in the same two-digit industry. Indeed, even with detailed four-digit industry dummies, the estimated φ_p vector shows substantial widening in the distribution of earnings among establishments.

V. Establishment Earnings and Labor Productivity

Was the increased dispersion of earnings among establishments accompanied by increased dispersion of other measures of establishment performance, or was the widening of the distribution unique to earnings? It would be strange if earnings were the only establishment-level factor that diverged among establishments. Divergence of earnings due to the labor market factors would presumably lead establishments with increasing wages to substitute other factors for labor—capital or innovative technology—and raise labor productivity relative to establishments with decreasing earnings, producing a wider dispersion of labor productivity. In terms of exit or entry of establishments, those with low productivity would presumably survive better in a world where they can hire workers at wages far below average than if wages are concentrated near the average,[21] further widening the distribution of productivity. From the productivity side, establishments in

[20] From 1977 to 2007, the mean number of employees in establishments increased from 18.4 to 20.0, but the standard deviation fell from 150 to 140. The mean number of employees in MU firms increased from 251.6 to 374.5, driven by increases in establishments per firm from 5.8 to 9.4; but the MU share of employment held fixed at 54% (based on LBD computations for 3,685,505 establishments in 1977 and 6,196,382 establishments in 2007).

[21] Grout (1984), Moene and Wallerstein (1997), and Acemoglu (2003) examine how earnings differentials and rent-sharing affects incentives to invest and implement new technology. Freeman and Kleiner (2005) show how different wage-setting policies influenced the exit pattern of plants in the declining shoe industry.

markets with inherent heterogeneity in workplace productivity (see, e.g., Melitz 2003; Klette and Kortum 2004; Faggio, Salvanes, and Van Reenen 2007; Haltiwanger, Lane, and Spletzer 2007; Andersson et al. 2008; Bender et al. 2008; Comin, Groshen, and Rabin 2009) due to differences in the introduction of new technology or other supply shocks or that face differential changes in product demand are likely to see productivity increases spilling over to wages through "rent-sharing" behavior. Efficiency wage models focused on the motivational impact of wages also suggest that wages and productivity are likely to increase or decrease together. Whatever the causal mechanism, we expect rising dispersion in earnings to be associated with rising dispersion of labor productivity.

To examine the relation between changes in productivity and changes in establishment earnings, we estimated the variance of log revenues per worker among establishments and compared changes in that variance to changes in the variance of log earnings. To estimate establishment revenues per worker, we obtained data from the US Census Bureau's Economic Census files, which are based on quinquennial censuses of establishments in every year with an ending of 2 or 7.[22] Revenues per worker are not an ideal indicator of productivity, but they have the virtue of focusing on the flow of funds that is likely to bound labor payrolls.

Table 6 displays the variance of log revenues per worker (upper panel) and the corresponding variance of log earnings among establishments (lower panel) from the LBD for the one-digit private sector industries every 5 years from 1977 to 2007. The variances of log revenues per worker are much larger than the variances of log yearly earnings—2–3 times larger for all sectors. More important for our issue, the variances of log revenues per worker increased roughly twice as much as the variances of log earnings (0.311 vs. 0.156). For whatever reason, establishments moved further apart in revenue per worker than they did in earnings in the period under study.

Rent-sharing and other noncompetitive models of wage determination offer one possible explanation for the two series diverging in the same period. These models posit that exogenous changes in revenues/profits change wages in the same direction (see, e.g., Margolis and Salvanes 2001; Arai 2003; Martins 2009; Dobbelaere and Mairesse 2010; Card, Heining, and Kline 2013). Following this logic, we examine the link between wages and revenues using the following model:

$$\ln w_{pir} = a + b\ln R_{pir} + c\ln \mathrm{AW}_{ir} + ds_I + v_{pir}, \tag{7}$$

where R_{pir} is revenue per worker in establishment p, industry i, and region r; AW_{ir} is the average wage of industry i in region r, an indicator of alter-

[22] See http://factfinder2.census.gov/faces/nav/jsf/pages/programs.xhtml?program=econ.

Table 6
Variance of Revenues per Worker and Earnings per Worker, 1977–2007

	1977	1982	1987	1992	1997	2002	2007	Change, 1977–2007
Variance of log revenues per worker:								
All sectors	.954	.965	.949	1.020	1.113	1.126	1.265	.311
Mining, Utilities, and Transportation	.421	.463	.670	.821	.860	.827	.967	.546
Manufacturing	.593	.633	.638	.656	.686	.646	.742	.149
Trade	1.135	1.129	1.115	1.165	1.228	1.207	1.280	.145
Finance, Insurance, and Real Estate	.911	.917	1.222	1.075	1.244	1.190	1.432	.521
Personal Services	.444	.426	.471	.459	.531	.565	.593	.149
Business Services	.878	.852	.914	.923	1.083	1.089	1.116	.238
Communication	.444	.430	.522	.748	.718	.736	.854	.410
Health, Education, and Social Services	.316	.559	.390	.402	.448	.567	.534	.218
Variance of log earnings:								
All sectors	.332	.362	.412	.413	.443	.446	.488	.156
Mining, Utilities, and Transportation	.302	.317	.328	.327	.323	.313	.316	.014
Manufacturing	.187	.204	.220	.218	.234	.226	.239	.052
Trade	.340	.353	.388	.390	.415	.413	.423	.083
Finance, Education, and Social Sciences	.202	.303	.433	.447	.467	.516	.579	.377
Personal Services	.364	.386	.408	.296	.321	.338	.370	.006
Business Services	.478	.506	.551	.547	.581	.582	.634	.156
Communication	.214	.269	.299	.355	.383	.474	.485	.271
Health, Education, and Social Services	.247	.229	.262	.249	.249	.236	.270	.023

NOTE.—Log of revenues per worker is taken from the Economic Census. Log of earnings is taken from the Longitudinal Business Data base. Figures for all sectors from the Economic Census are based on the sectors available in the table every census year. The Economic Census expanded in scope over the 1977–2002 period, but the Business Register and LBD covered all industries throughout. As a check, we calculated the variance of revenues per worker restricted to industries where in each year total industry employment in the economic census is greater or equal to 90% of total industry employment in the LBD. The variance trend is very similar (for 1977, the variance is 0.945; for 1982, it is 0.965; for 1987, it is 0.991; for 1992, it is 1.036; and for 1997, it is 1.111), where the difference is calculated from the first available year in the table.

native wages that would affect w_{pir} in the establishment and region through supply conditions; and s_I is a composite index of observable characteristics measured at the detailed industry level.[23]

Table 7 presents the results from estimating equation (7) on a panel of establishments for 5-year intervals from 1977 to 2007 using different statistical models. The key coefficient in the regression is the b parameter that

[23] The skills measure is the average predicted xb from the Sec. II equations using the yearly CPS files, where x includes education, experience and its square, all interacted with gender. We averaged the skill measure by detailed industry using the definition of ind50 from the IPUMs to match each year to sic3 and naics4.

Table 7
Establishment Wage Regressions

	Dependent Variable: Log Earnings (Establishment Wage)		
		Fixed Establishment Effects	
	OLS	FE	IV-FE
Log earnings (sales/employees)	.386	.324	.163
	(.000)	(.000)	(.002)
Individuals' predicted wage in industry:			
Log earnings (predicted industry wage)	.553	.051	.062
	(.001)	(.002)	(.002)
Alternative wage:			
Log earnings (industry × region average)	.343	.113	.131
	(.001)	(.001)	(.002)
1-digit industry controls	Yes	No	No
Fixed establishment effects	No	Yes	Yes

NOTE.—N = 7,188,373. The model is estimated on a panel of establishments from 1977 to 2007, quinquennial observations from the Economic Census. The models include controls for observation year and establishment age. Predicted industry wage is calculated from a log earnings equation including years of education, experience, experience squared, interacted with gender, averaged at the industry level using yearly CPS data. Instrumental variable (IV) specifications use industry revenue per worker, averaged over all regions except own region, as the instrument for revenue per worker.

links revenues per worker to earnings. Given the fact that the variance of revenues per worker increased at about twice the variance of earnings, an estimated b of around 0.7 would attribute most of the increased variance of log earnings to the increased variance of revenues per worker.[24] None of our estimated models give such a large rent-sharing parameter. The ordinary least squares model in column 1 has a rent-sharing parameter of 0.386. Addition of establishment fixed effects in column 2 (so that the analysis links within establishment earnings to within-establishment revenues per worker) drops the rent-sharing parameter to 0.324. The instrumental variable estimate in column 3, which deals with the endogeneity of revenues per worker by the Card, Devicienti, and Maida (2010) method of taking revenues outside of the region of the observed establishment as the instrument, gives an estimate of 0.163. The identifying restriction in this analysis is that, conditional on average earnings in the industry and region, higher revenues per worker in the industry outside the region affect earnings solely through establishment revenues. With an impact on earnings of 0.163, the increased revenue per worker adds about 5%–6% to the variance of earnings among establishments and thus falls far short of explain-

[24] The variance (var) decomposition of (7) links ΔVar log earnings to $b^2\Delta$Var log(revenues per worker), all else the same. With ΔVar log(revenues per worker) about twice the magnitude of ΔVar log earnings, $b \sim .7$ would give the $b^2 \sim 1/2$ necessary for the changed variance in revenues to account for the changed variance in earnings.

ing the increased variance of establishment effects and increased inequality of individual earnings.[25]

There are caveats to the finding that rent-sharing does not appear to be a major driver of the increased dispersion of establishment level earnings. Our measure of revenue per worker is a serviceable but not ideal measure of productivity nor of the supernormal profits that firms may share with workers. Our measure of establishment pay may be contaminated by changes in the composition of workers. The importance of rent-sharing may differ by the industry in which an establishment operates and the industrial organization of the sector or by the firm to which an establishment belongs. Given that approximately 43% of American workers are covered by "shared capitalist" forms of compensation, such as profit-sharing and gain-sharing systems, use of bonuses, stock options, and employee-share ownership (Kruse, Blasi, and Park 2010, table 1.1), more research on rent-sharing is warranted.[26] Still, our results suggest that factors beyond demand-driven rent-sharing would seem to be needed to account fully for the divergence of establishments in earnings space.

VI. Conclusion

The distribution of earnings across establishments widened markedly during the 1970s to 2000s period of increasing inequality of individual earnings. Using several data sets and modeling procedures, we find that the widening of the establishment earnings distribution underlies much of the increase in inequality. The widening establishment distribution accounted for most of the increased variance of log earnings among workers, explaining 79% of the increase in dispersion among workers who continued from one year to the next in the same establishments. It also accounted for most of the pattern of larger increases in earnings among workers higher in the earnings distribution and for most of the increased gap between earners in the upper 5% and others. The distribution of log revenues per worker also widened over the period, although our demand-driven rent-sharing model did not add much to the change in variance of earnings.

The finding that the establishments where people work have been a major factor in the much-heralded increase in inequality does not provide a

[25] Assuming that the variance of revenues per worker increased by twice the increased variance in log earnings, the contribution of the increase in revenues per workers would be $(.163)^2(2) = .054$.

[26] Studies that link changes in dispersion of wages to modes of compensation give conflicting results for the contribution of compensation systems to the increased dispersion of pay among individuals. Lemieux, MacLeod, and Parent (2009) find that an increase in the incidence of performance pay explains 24% of the increase in individual wage dispersion from the late 1970s to the early 1990s. But Gittleman and Pierce (2015) report that the incidence of performance pay declined markedly since the mid-1990s and that performance pay contributed only modestly to the growth in wage variance between 1994 and 2010.

simple explanation for the increase in inequality but rather directs attention at the economics that have pulled establishments apart in earnings space, which preceding work has downplayed. This, in turn, suggests the value of renewed analysis of establishment pay-setting and hiring policies on the demand side, on establishment-level mobility of labor on the supply side, and on factors beyond establishment demand shocks, such as productivity shocks associated with the introduction of innovative products or processes, in producing the divergence of establishment earnings. The huge role of establishment factors in the trend rise in inequality found in this study is a signpost to pay attention to the places where people work as well as to their skills in studies of inequality.

Appendix

CPS March Data

We use the internal Census March Current Population Survey from survey years 1978–2010 to obtain observations of weekly wages from 1977 to 2009. All samples include workers of ages 16–64 with more than 5 hours per week last year, more than 12 weeks worked last year, and whose class of work in their longest job last year was private or government wage/salary employment. Students, agricultural workers, public administration workers, and those in the armed forces are excluded. Weekly earnings are calculated as annual earnings divided by the weeks worked in the prior year. Gross earnings include wages, salaries, overtime, tips, and commissions. Allocated earnings observations are excluded using the earnings allocation flags. Educational codes are transformed to grade levels using Jaeger (1997) and subsequent adaptations. Final weights are used in all calculations. Observations with a real wage below half the minimum wage level in 1982 were excluded.

US Census Bureau's Longitudinal Business Database (LBD)

The establishment level data are from the US Census Bureau's Longitudinal Business Database (LBD) covering the period 1977–2009. The data include all private employers for all sectors except agriculture. The data are sourced from the Census Bureau's Business Register, which is continually updated using administrative records, Economic Census returns, and surveys such as the Company Organization Survey. The file collects establishment payroll and employment data, which is used to calculate the average establishment wage per worker for observations with positive employment and payroll.

Survey respondents are asked to follow the definition of salaries and wages used for calculating the federal withholding tax. They report the gross earnings paid in the calendar year to employees at the establishment prior to such deductions as employees' social security contributions, withhold-

ing taxes, group insurance premiums, union dues, and savings bonds. Included in gross earnings are all forms of compensation, such as salaries, wages, commissions, dismissal pay, paid bonuses, vacation and sick leave pay, and the cash equivalent of compensation paid in kind. Salaries of officers of the establishment, if a corporation, are included. Payments to proprietors or partners, if an unincorporated concern, are excluded. Salaries and wages do not include supplementary labor costs such as employers' social security contributions and other legally required expenditures or payments for voluntary programs. The definition of payrolls is identical to that recommended to all federal statistical agencies by the Office of Management and Budget.

Wages are converted to constant 2002 dollars using the Consumer Price Index. Establishments are excluded that have an average wage less than half the yearly equivalent of the 1982 minimum wage of $3.35 an hour (CPI deflated) for a 40-hour week. Establishments with over 100,000 employees are also excluded, as from observation these are generally firm-level or miscoded records, and we are not aware of a US establishment that large. One issue with our wage measure is that payroll is reported annually and employment is reported for the week of March 12. The establishment wage can be affected by significant changes in establishment employment within the year.

The LBD follows establishments over time, where considerable effort was invested by the US Census to recover longitudinal identifiers through linking records and matching names and addresses (Jarmin and Miranda 2002). We use these identifiers to define establishment births, deaths, and continuers. A *birth* is an establishment that is observed in the data that did not exist 5 years earlier. Similarly, a *death* is an observation that does not survive the 5 years until the next economic census year. Establishments are either single-unit (SU) establishments, where the (generally smaller) firm produces in one location, or multi-unit (MU) establishments that are part of a company that operates at multiple locations. The 10%, median, and 90% deciles are calculated by taking a neighborhood of establishments 1% on either side of the decile and using the mean of this sample as a pseudo-decile.

LEHD 1992–2007 and Matched LEHD Data

The LEHD data, produced by the LEHD program at the US Census Bureau, consists of annualized quarterly earnings from the Unemployment Insurance (UI) benefit programs, linked to the Quarterly Census of Employment and Wages Program. The construction of the LEHD data is carefully described in Abowd et al. (2002). We use data from the nine states that were part of this program from 1992 onward. The nine states are California, Colorado, Idaho, Illinois, Maryland, North Carolina, Oregon, Washington, and Wisconsin. They cover half of the US employment. For robust-

ness, we also examined results for 16 states that were part of the program since 1997 and 22 states that were part of the program from 2002, and we found very similar results (results are available from the authors on request). Our LEHD sample is limited to workers who had a job in all four quarters of a year, who were between 16 and 64 years of age, and whose quarterly earnings were larger than the level of a full-time job with more than half the minimum wage in 1982. For robustness, we also report some results from a sample limited to those with a valid job in the second quarter only (not requiring full-year employment).

Employers are identified by state unemployment insurance account number (SEIN), and establishments are identified as a specific place of work (SEIN UNIT). According to Abowd et al. (2014), 70% of employment occurs in firms with only a single establishment within a state. Among multiplant employers within the same state, establishment is allocated by the LEHD program, using the distance between place of residence and place of work, conditional on summing up restrictions (see Abowd et al. [2002] for details). We report a robustness check using SEIN rather than the more detailed SEIN UNIT as our unit of analysis.

Because education is imputed in the LEHD data, we have added data on education with a match between the LEHD and the US 2000 and 1990 Census long form data and the CPS when available (1986–92). We first match the records between the Economic History files (EHF) of the LEHD and the 2000 Census. For those who do not match, we merge in data from the 1990 Census, and for those who do not match any of these, we merge in information on education from the Current Population Survey. The Census long form covers approximately 15% of the population in 2000, and using this procedure, we are able to create a sample covering 18% of the jobs in the EHF data from 1992, 1997, 2002, and 2007. This data set is labeled "matched LEHD sample" in the analysis above.

Table A1
Correcting Establishment Earnings Dispersion Using LEHD

	1992	2007	1992–2007
Variance (LBD average log earnings)	.412	.487	.075
Correcting LBD comparison using LEHD data:			
$\text{Cov}(\mu_f, \sigma_f)$	.009	.019	
$\text{Var}(\sigma_f)$	.051	.048	
$(1/4)[V(\sigma_f) + \text{Cov}(\mu_f, \sigma_f)]$	.026	.031	
Var (ln$w$) corrected	.386	.456	.070

NOTE.—The US Census Bureau's Longitudinal Business Database (LBD) earnings are average wages per worker (annual earnings divided by March 12 employment). Variable LEHD earnings is based on annual earnings for full-year employees from the nine 1992 LEHD states, from quarterly earnings from the Unemployment Insurance files, and within-establishment dispersion and means are calculated within SEIN-UNIT per state year using Economic History files (EHF) and then aggregated, matched, and disaggregated to the appropriate LBD-Units in the LBD files. Establishment figures are employment weighted.

Table A2
Dispersion Measures of the Current Population Survey and the Longitudinal Business Database, 1977–2009

	1977	1982	1987	1992	1997	2002	2007	2009	Growth
Weekly earnings CPS:									
V(ln earnings)	.463	.501	.524	.546	.574	.601	.592	.633	.170
Predicted	.206	.207	.199	.209	.214	.222	.215	.229	.023
Residual	.257	.294	.324	.337	.36	.379	.377	.404	.147
Weekly earnings LEHD states, CPS:									
V(ln earnings)				.538	.553	.589	.618	.623	.085
Predicted				.202	.199	.209	.222	.220	.017
Residual				.336	.355	.380	.396	.403	.067
Hourly wage, CPS:									
V(ln wage)	.306	.323	.352	.369	.391	.426	.425	.443	.137
Predicted	.118	.119	.121	.122	.128	.139	.140	.141	.023
Residual	.	.209	.231	.247	.264	.287	.284	.302	.114
Weekly relative earnings, CPS:									
Decile ratio d9/d5	2.06	2.20	2.28	2.36	2.40	2.52	2.52	2.65	.58
Decile ratio d5/d1	2.55	2.69	2.67	2.62	2.56	2,57	2.63	2.81	.27
Average relative establishment earnings LBD:									
Decile ratio d9/d5	2.22	2.46	2.55	2.61	2.64	2.61	2.65		.44
Decile ratio d5/d1	2.30	2.35	2.38	2.29	2.39	2.44	2.52		.22
Gini index:									
CPS individuals	.357	.365.	.370	.398	.431	.443	.446		.089
LBD establishments	.299	.320	.342	.346	.365	.372	.393		.094

NOTE.—Sample of all wage earners ages 16–64; see the appendix for details. Weekly earnings are earnings last year divided by weeks worked last year. Hourly wage is weekly earnings divided by usual number of hours per week. The LEHD states are the nine states in the LEHD data from 1992 onward. LBD numbers are employment weighted.

Table A3
Variance Decomposition of LEHD Earnings with Individual Characteristics: Within and Between Employer (SEIN)

	1992	1997	2002	2007	1992–2007 Change	Share of Change
All:						
Var (ln w)	.457	.478	.500	.545	.088	1.00
Skills: Var(s)	.110	.104	.103	.102	−.008	−.08
Worker-worker: ρ	.316	.310	.311	.322	.006	
Worker-establishment: ρ_{φ}	.221	.229	.246	.287	.065	
Variance between	.222	.231	.244	.276	.054	.61
Establishment effect: $V(\varphi)$	.139	.152	.161	.184	.046	.52
Skills contribution: $V(s)\rho$	.035	.032	.032	.033	−.002	−.02
Match contribution: $V(s)2\rho_{\varphi}$	.049	.047	.051	.059	.010	.11
Variance within	.236	.246	.257	.270	.034	.39
Within residual $V(u)$	.160	.175	.186	.200	.040	.45
Skills contribution: $V(s)(1 - \rho)$	.075	.071	.071	.069	−.006	−.07
No. of individuals (millions)	3.9	4.2	4.3	4.3		
No. of employers (SEIN; millions)	.6	.6	.6	.7		‘

NOTE.—The table is identical to table 3, but it uses employer (SEIN) instead of establishment (SEIN UNIT). Matched LEHD Data. See the appendix for details.

References

Abowd, John, John Haltiwanger, Julia Lane, Kevin L. McKinney, and Kristin Sandusky. 2007. Technology and the demand for skill: An analysis of within and between firm differences. NBER Working Paper no. 13043 (April), National Bureau of Economic Research, Cambridge, MA.

Abowd, John M., Francis Kramarz, and David N. Margolis. 1999. High wage workers and high wage firms. *Econometrica* 67, no. 22:251–334.

Abowd, John M., Francis Kramarz, Sébastien Pérez-Duarte, and Ian M. Schmutte, 2014. Sorting between and within industries: A testable model of assortative matching. NBER Working Paper no. 20472, National Bureau of Economic Research, Cambridge, MA.

Abowd, John M., Paul A. Lengermann, and Kevin L. McKinney. 2003. The measurement of human capital in the U.S. economy. Technical Paper TP-2002-09 (March), US Census Bureau, Washington, DC.

Abowd, John M., Bryce E. Stephens, Lars Vilhuber, Fredrik Andersson, Kevin L. McKinney, Marc Roemer, and Simon Woodcock. 2002. The LEHD infrastructure files and the creation of the Quarterly Workforce Indicators. Longitudinal Employer-Household Dynamics Technical Papers 2002–05, Center for Economic Studies, US Census Bureau, Washington, DC.

Acemoglu, Daron. 2003. Labor- and capital-augmenting technical change. *Journal of European Economic Association* 1, no. 1:1–37.

Aitchison, J., and J. A. C. Brown. 1957. *The lognormal distribution*. Cambridge: Cambridge University Press.

Alvaredo, Facundo, Anthony B. Atkinson, Thomas Piketty, and Emmanuel Saez, 2013. The top 1 percent in international and historical perspective. *Journal of Economic Perspectives* 27, no. 3:3–20.

Andersson, Fredrik, Clair Brown, Benjamin Campbell, Hyowook Chiang, and Yooki Park. 2008. The effects of HRM practices and R&D investment on worker productivity. In *The analysis of firms and employees: Quantitative and qualitative approaches*, ed. Stefan Bender, Julia Lane, Kathryn L. Shaw, Fredrik Andersson, and Till von Wachter. Chicago: University of Chicago Press.

Arai, Mahmood. 2003. Wages, profits, and capital intensity: Evidence from matched worker-firm data. *Journal of Labor Economics* 21, no. 3:593–618.

Autor, David H., Lawrence F. Katz, and Melissa S. Kearney. 2008. Trends in U.S. wage inequality: Revising the revisionists. *Review of Economics and Statistics* 90, no. 2:300–323.

Bell, Linda, and Richard Freeman. 1991. The causes of rising interindustry wage dispersion in the United States. *Industrial and Labor Relations Review* 44, no. 2:275–87.

Bender, Stefan, Julia Lane, Kathryn Shaw, Fredrik Andersson, and Till von Wachter, eds. 2008. *The analysis of firms and employees: Quantitative and qualitative approaches*. Chicago: University of Chicago Press.

Burkhauser, Richard, Shuaizhang Feng, Stephen Jenkins, and Jeff Larrimore. 2011. Estimating trends in U.S. income inequality using the Current Population Survey: The importance of controlling for censoring. *Journal of Economic Inequality* 9, no. 3:393–415.

Card, David, Fransisco Devicienti, and Agata Maida. 2010. Rent-sharing, holdup, and wages: Evidence from matched panel data. NBER Working Paper no. 16192 (July), National Bureau of Economic Research, Cambridge, MA.

Card, David, Jörg Heining, and Patrick Kline. 2013. Workplace heterogeneity and the rise of West German wage inequality. *Quarterly Journal of Economics* 128, no. 3:967–1015.

Comin, Diego, Erica L. Groshen, and Bess Rabin. 2009. Turbulent firms, turbulent wages? *Journal of Monetary Economics* 56, no. 1:109–33.

Davis, Steve J., and John Haltiwanger. 1991. Wage dispersion between and within U.S. manufacturing plants, 1963–86. *Brookings Papers on Economic Activity: Microeconomics*, no. 2:115–200.

Dickens, William T., and Lawrence F. Katz. 1987. Inter-industry wage differences and industry characteristics. In *Unemployment and the structure of labor markets*, ed. Kevin Kevin and Jonathan Leonard. Oxford: Basil Blackwell.

Dobbelaere, Sabien, and Jacques Mairesse. 2010. Micro-evidence on rent sharing from different perspectives. NBER Working Paper no. 16220 (July), National Bureau of Economic Research, Cambridge, MA.

Dunne, Timothy, Lucia Foster, John Haltiwanger, and Kenneth R. Troske. 2004. Wage and productivity dispersion in US manufacturing: The role of computer investment. *Journal of Labor Economics* 22, no. 2:397–430.

Faggio, Giulia, Kjell Salvanes, and John Van Reenen. 2007. The evolution of inequality in productivity and wages: Panel data evidence. NBER Working Paper no. 13351 (August), National Bureau of Economic Research, Cambridge, MA.

Freeman, Richard B., and Morris M. Kleiner. 2005. The last American shoe manufacturers: Decreasing productivity and increasing profits in the shift from piece rates to continuous flow production. *Industrial Relations* 44, no. 2:307–30.

Gibbons, Robert, and Lawrence Katz. 1992. Does unmeasured ability explain interindustry wage differentials? *Review of Economic Studies* 59, no. 3 (July): 515–35.

Gittleman, Maury, and Pierce Brooks. 2015. Pay for performance and compensation inequality: Evidence from the ECEC. *Industrial and Labor Relations Review* 68, no. 1 (January): 28–52.

Goldin, Claudia, and Lawrence Katz. 2008. *The race between education and technology*. Cambridge, MA: Harvard University Press.

Groschen, Erica. 1991. Sources of intra-industry wage dispersion: How much do employers matter? *Quarterly Journal of Economics* 106, no. 3: 869–84.

Grout, Paul A. 1984. Investment and wages in the absence of binding contracts: A Nash bargaining approach. *Econometrica* 52, no. 2 (March): 449–60.

Gruetter, Max, and Rafael LaLive. 2009. The importance of firms in wage setting. *Labour Economics* 16, no. 2:149–60.

Haltiwanger, John C., Julia I. Lane, and James R. Speltzer. 2007. Wages, productivity, and the dynamic interaction of businesses and workers. *Labour Economics* 14, no. 3:575–602.

Hellerstein, Judith K., David Neumark, and Kenneth R. Troske. 1999. Wages, productivity, and worker characteristics: Evidence from establishment-level production functions and wage equations. *Journal of Labor Economics* 178, no. 3:409–46.

Jaeger, David A. 1997. Reconciling educational attainment questions in the CPS and the Census. *Monthly Labor Review* 120, no. 8:36–40.

Jarmin, Ron S., and Javier Miranda. 2002. The longitudinal business database. Working Paper no. 02-17, Center for Economic Studies, Washington, DC.

Klette, Tor Jakob, and Samuel Kortum. 2004. Innovating firms and aggregate innovation. *Journal of Political Economy* 112, no. 5:986–1018.

Kremer, Michael, and Eric Maskin. 1996. Wage inequality and segregation by skill. NBER Working Paper no. 5718, National Bureau of Economic Research, Cambridge, MA.

Krueger, Alan B., and Lawrence H. Summers. 1988. Efficiency wages and the inter-industry wage structure. *Econometrica* 56, no. 2:259–94.
Kruse, Douglas L., Joseph Blasi, and Rhokeun Park. 2010. Shared capitalism in the U.S. economy: Prevalence, characteristics, and employee views of financial participation in enterprises. In *Shared capitalism at work: Employee ownership, profit and gain sharing, and broad-based stock options*, ed. Douglas L. Kruse, Richard B. Freeman, and Joseph R. Blasi. Chicago: University of Chicago Press.
Lane, Julia I., Laurie A. Salmon, and James R. Spletzer. 2007. Establishment wage differentials. *Monthly Labor Review* 130, no. 4:3–17.
Lazear, Edward, and Kathryn Shaw, eds. 2009 *The structure of wages: An international comparison*. Chicago: University of Chicago Press.
Lemieux, Thomas. 2006. Increasing residual wage inequality: Composition effects, noisy data, or rising demand for skill? *American Economic Review* 96, no. 3:461–98.
———. 2008. The changing nature of wage inequality. *Journal of Population Economics* 21, no. 1:21–48.
Lemieux, Thomas, W. Bentley MacLeod, and Daniel Parent. 2009. Performance pay and wage inequality. *Quarterly Journal of Economics* 134, no. 1:1–49.
Margolis, David N., and Kjell G. Salvanes. 2001. Do firms really share rents with their workers? IZA Discussion Paper no. 330 (July). Institute for the Study of Labor, Bonn, Germany.
Martins, Pedro. 2009. Rent sharing before and after the wage bill. *Applied Economics* 41, no. 17:2133–51.
Melitz, Marc J. 2003. The impact of trade on intra-industry reallocations and aggregate industry productivity. *Econometrica* 71, no. 6:1695–1725.
Mincer, Jacob. 1974. *Schooling, experience, and earnings*. New York: National Bureau of Economic Research.
Moene, Karl Ove, and Michael Wallerstein. 1997. Pay inequality. *Journal of Labor Economics* 15, no. 3:403–30.
Nordstöm Skans, Oskar, Per-Anders Edin, and Bertil Holmlund. 2009. Wage dispersion within and between plants: Sweden, 1985–2000. In *The structure of wages: An international comparison*, ed. Edward P. Lazear and Kathryn L. Shaw. Chicago: University of Chicago Press.