Augmenting the Human Capital Earnings Equation with Measures of Where People Work

Abstract

We augment standard ln earnings equations with variables reflecting unmeasured attributes of workers and measured and unmeasured attributes of their employer. Using panel employee-establishment data for US manufacturing we find that the observable employer characteristics that most impact earnings are: number of workers, education of co-workers, capital equipment per worker, industry in which the establishment produces, and R&D intensity of the firm. Employer fixed effects also contribute to the variance of ln earnings, though substantially less than individual fixed effects. In addition to accounting for some of the variance in earnings, the observed and unobserved measures of employers mediate the estimated effects of individual characteristics on earnings and increasing earnings inequality through the sorting of workers among establishments.

Erling Barth, Institute for social research, ESOP, University of Oslo, and NBER
James Davis, US Bureau of the Census and BRDC
Richard B. Freeman, Harvard and NBER
August 6, 2016

We have benefited from support from the Labor and Worklife Program at Harvard University, NBER and from the Norwegian Research Council (projects # 202647 and 199836 (Barth)). Thanks to Thomas Lemieux for very useful comments. Any opinions and conclusions expressed herein are those of the authors and do not necessarily represent the views of the U.S. Census Bureau. All results have been reviewed to ensure that no confidential information is disclosed.

Standard earnings equations relate ln earnings to the human capital/demographic attributes[1] of individuals. While the standard equation accounts for a sizable proportion of the distribution of individual earnings, there is a sufficiently wide and increasing dispersion of earnings among workers with the same measured characteristics to challenge "the law of one price" in the US labor market. Exemplifying the wide dispersion of earnings in the US for workers with similar skills, Devroye and Freeman (2001) found that variance of ln earnings among US workers within narrow bands of adult literacy test scores exceeded the variance of earnings among **all workers** in the United Kingdom, the Netherlands, and Sweden.

To what extent do the measured and unmeasured characteristics of employers contribute to the variation in earnings among similarly skilled workers? Does taking account of the characteristics of employers alter estimates of how individual characteristics affect earnings? To what extent do employees with similar attributes work together? To what extent do high/low wage firms hire workers with similar attributes?

This paper seeks to answer these questions about wage determination and the allocation of labor among high and low paying employers by augmenting the standard log earnings equation with attributes of employers in US manufacturing[2]. It combines earnings data for individuals from the Longitudinal Employer Household Dynamics (LEHD) with data on worker attributes from the Decennial Census and the CPS, data on establishments from the Census of Manufacturing, and data on firms from the Longitudinal Business Database (LBD) and the Survey of Industrial Research and Development (SIRD)[3] It uses the observed and unobserved components of the augmented earnings regression to assess the extent to which the labor market sorts workers into establishments with similar workers or between high and low paying establishments.[4]

---

1 We use "attributes" and "characteristics" interchangeably in this paper.

2 Previous studies of the importance of the employer for wage determination include Davis and Haltiwanger (1991), Groschen (1991), Abowd, Kramarz and Margolis (1999), Lane, Salmon, and Spletzer (2007), Gruetter and LaLive (2009), Nordström Skans, Edin, and Holmlund (2009), Card, Devicienti and Maida (2010), Barth et al. (2016), and Song et al. (2015). The literature on rent sharing relates individual earnings to the productivity or profitability of the establishment, see eg. Hellerstein, Neumark and Troske (1999), Margolis and Salvanes (2001), Dunne, Foster, Haltiwanger and Troske (2004), Faggio, Salvanes and van Reenen (2007), Dobbelaere and Mairesse (2010), Mortensen, Christensen, and Bagger (2010), and Card, Heining, and Kline (2013).

3 These data measure: firm employment, establishment employment, capital per worker, the percentage of output exported overseas, and R&D per employee and to estimate the average characteristics of the establishment work force: years of schooling, age, gender and race.

4 Following Abowd, Kramarz and Margolis (1999), several analysts have examined these questions, including Andrews, Schank, and Upward (2008), Mendes, van den Berg, and Lindeboom (2010), Lise, Meghir, and Robin (2013), Card et al. (2013), and Abowd et al. (2014) Our measures of years of schooling and of time varying establishment characteristics provides a different take on sorting along both dimensions.

Methodology

The traditional cross section human capital wage equation (Mincer 1974) relates earnings w_{it} of individual i in period t to observable measures/indicators of personal skill and other individual characteristics that ideally reflect productivity but that also reflect employer attitudes or perceptions resulting from prejudicial or statistical discrimination:

(1) $ln\ w_{it} = \beta_0 + \gamma_t + x_{it}\beta + u_{it}$

where γ_t is a vector of time dummies, x_{it} is a vector that includes years of schooling and individual attributes such as age, gender and race. The equation does not include attributes of the establishment or firm, though they can be added to reflect compensating differential or other factors related to the full compensation of workers that are not captured by the earnings measure.

Our augmented earnings equation adds to equation (1) the measured and unmeasured characteristics of an individual's establishment/firm:

(2) $\ln w_{ijt} = \beta_0 + \gamma_t + x_{it}\beta + z_{jt}\,d + \psi_{ij} + e_{ijt}$

where j(i) is an index of the workplace which employs individual i at time t, and where ψ_{ij} is a unique job/workplace fixed effect for every individual and workplace pair. The t subscript on z_{jt} allows observed employer characteristics to vary over time also within each job. If workplace j has a capital stock K, it is assumed that K affects the wages of all workers similarly. With a panel of workers and employers, the d coefficients for the establishment characteristics are estimable using within job variation in the z. For example, if z relates to employment (larger establishments pay more), the effect of employment on wages can be estimated for the same worker/job when the establishment changes employment.

Multiple observations on a single person, including employer identifiers in longitudinal data, allow us to decompose the job/employer effect into an individual fixed effect measured by the coefficient on a dummy variable for an individual, an establishment fixed effect, and a match component orthogonal to the individual and establishment fixed effects, $\psi_{ij} = \alpha_i + \phi_j + \xi_{ij}$, per the "AKM decomposition" (Abowd, et al. 1999). We further define: $\alpha_i = X_i B + a_i$, and $\phi_j = Z_j D + \varphi_j$ where X and Z are covariates fixed for each individual and establishment. We identify the B and D vectors of parameters by assuming that the residual of the individual fixed effect is orthogonal to individual fixed characteristics, and that the residual of the establishment fixed effect is orthogonal to establishment fixed characteristics. But our analysis allows the components of both fixed effects to be correlated with the time varying characteristics as well as to co-vary with each other. Our final augmented equation is:

(3) $lnw_{ijt} = \beta_0 + \gamma_t + x_{it}\beta + X_i B + a_i + z_{jt}\,d + Z_j D + \varphi_j + \xi_{ij} + e_{ijt}$
$= \beta_0 + \gamma_t + \omega_{it} + \Omega_{jt} + \xi_{ij} + e_{ijt}$

where ω_{it} is the individual component of the wage, and Ω_{jt} is the establishment component, both of which contain observable and unobservable parts.

Because the LEHD has no measures of education for workers[5], we matched workers on the LEHD to their attributes on the Census/CPS to measure their years of education, with a match rate of about 20 percent of all observations. To make maximum use of the full data set, however, we use the full sample to differentiate the establishment effects from the individual effects and use age and age square to measure experience. Appendix table A gives summary statistics for the matched and full sample. The matching process produces a matched sample that is higher in earnings and worker attributes that are positively associated with earnings such as age and being white and that is also higher in firm and establishment attributes positively associated with earnings such as number of employees and capital per employee. But the variation in characteristics is still large enough for our empirical analysis to yield meaningful statistical relations.

Comparing equations 1 and 3, if personal skills and attributes are the sole factor behind differences in earnings, the coefficients of equation (1) estimate the gross return to those skills/attributes *inclusive of possible gains from access to different employers,* while equation (3) measures the net return exclusive of the earnings characteristics of employers. Alternatively, to the extent that the covariates in equation (1) are correlated with the equation (2) variables, the estimated coefficients of (1) can be viewed as biased estimates of the net effects of skills/attributes in equation (2).

Census-LEHD & BRB Matched Data

To estimate augmented ln earnings equation (3) we combined data files for workers, establishments, and firms in manufacturing. We focus on manufacturing because the Annual Survey of Manufacturers provides information on manufacturing establishments annually that is unavailable for other sectors. Our methodology can, however, be applied to other sectors using data from the quinquennial Economic Censuses and other sources.

The dependent variable is earnings for individual workers, obtained from the Longitudinal Employer-Household Dynamics (LEHD) Employment History Files for the nine states with LEHD data from 1992 through 2007.[6] The LEHD data are linked to the quinquennial Census of Manufacturers (CoM) for the four economic census years, 1992, 1997, 2002 and 2007, and to the Annual Survey of Manufacturers (ASM) in the intermediate years, using the LEHD Business Register Bridge (BRB) that links data at the firm level. LEHD establishments are linked by firm, detailed industry and county to CoM/ASM establishments. For single-unit firms and for plants of a firm not located in the same county as other plants of the firm in the same industry, the mapping from LEHD to CoM/ASM establishments is unique within detailed

5 Education data have been recently now added to the LEHD.

6 The LEHD data provides annualized quarterly earnings from the unemployment insurance (UI) benefit programs, linked to the Quarterly Census of Employment and Wages Program. We use only observations that include positive earnings in the second quarter of the year. Abowd et al. (2002) describe the construction of the LEHD data. The nine states are: California, Colorado, Idaho, Illinois, Maryland, North Carolina, Oregon, Washington, and Wisconsin. They cover approximately half of US employment. Comparisons with data for states that cover different time periods show that the nine state sample are reasonably representative (Barth et al. 2016).

industry and county. This is the vast majority of observations. But for plants of a firm in the same industry and county, the link is not one-to-one. For these establishments, we aggregate plant characteristics to the firm-industry-county level and link these measures to their workers.

We obtain measures of the years of schooling, occupation, age, race and gender of workers in the LEHD by linking workers to their characteristics in the 1990 and 2000 Decennial Census long form and March CPS files for 1986-1997. The Census Center for Administrative Records staff matched these data using the internal person identifier called a Protected Identification Key (PIK), which is the person identifier in the Census and CPS and in the LEHD. Beginning with 2000 decennial files have very high PIK match rates of 90-93% (Mulrow et al. 2011, Rastogi and O'Hara 2012). However the 1990 PIK is more limited due to the vintage of address files,[7] so that matching the Census/CPS data to the LEHD Employee History Files (EHF) provided us with data on years of schooling and other worker attributes for 20.5% of employees in the LEHD data.[8]

The quinquennial Census of Manufacturing and Annual Survey of Manufacturers provides production-related data on manufacturing establishments, which we add to the files on employees: the number of workers at establishments, and establishment-level capital equipment and building stock (as constructed by Foster, Grim, and Haltiwanger (2016) with perpetual inventory methods). We measure firm employment from the LBD, and whether a firm reports R&D expenditures and the amount from the Survey of Industrial R&D. Appendix table A gives summary statistics for our key variables.

Basic Variance Decomposition

As the first step in unpacking the impact of employers on earnings, we decomposed the variance of the ln earnings of individuals in the full LEHD data into the variance within establishments and the variance between establishments for the whole US economy and for the eight large sectors of the US economy listed in table 1. We regressed the ln earnings of individuals on establishment dummies separately for each year and take the variance among the establishment dummies as our estimated establishment effect. The remaining variance reflects earnings differences within an establishment.

Table 1 displays the variance and the share of variance associated between establishments, and the growth of the variance and its share between establishments. In the economy as a whole 48 percent of the variance of ln earnings among workers comes from variation between establishments, while 66 percent of the .091 *growth* in variance arose from a widening of the earnings distribution between establishments.[9] In manufacturing 57 percent of

7 Individual name and address files are highly sensitive and not generally distributed in the Census Bureau with the data files. Our versions of 1990 decennial files did not have original name and address data, and had to be reconstructed with other data sets. As a result, the PIK matches favor less mobile adult heads of household.

8 We first matched to the 2000 Census, then matched missing cases to the 1990 Census, and finally matched missing cases to the CPS data.

9 Though our sample differs from Barth et al. (2016) because we include all jobs observed in the 2nd quarter of each year, not only full year -, main jobs as in that paper, the results are consistent. We included all jobs to keep

the 0.092 growth in variance come from growth between establishments.

Table 2 turns to the subset of manufacturing workers for whom we match observations in the LEHD and Census of Manufacturers to Decennial Census or CPS files which have the following characteristics: 1) the CPS/Census has years of schooling; (2) the CoM reports establishment capital and firm and establishment employment[10]; and (3) where the workers are in our data at least four times. The variance of ln earnings in the matched table 2 panel is noticeably smaller than the variance of ln earnings for manufacturing in the full LEHD in table 1. We attribute this to the fact that the matched data has fewer small firms than the full sample (see Appendix table A) and to our requirement that a person be observed at least four times to be in the file, which drops transitory workers. In addition, the increase in variance falls short of the increase in the full LEHD; and the 43% contribution of increased earnings between establishments to the increase in inequality is smaller than the 57% in the full LEHD. The matched sample thus is likely to *understate* the contribution of establishments to the variation in earnings.

As establishments belong to firms that include other establishments, it is important to differentiate establishment effects from firm effects in analysis. To assess the importance of firms in the variance of earnings among establishments, we regressed the estimated establishment fixed effects on dummy variables for firms. The proportion of the variance attributed to firms reflects the overall pay practices of firms while the remaining proportion reflects pay differences among establishments in the same firm. These calculations show that 90.4% (= 0.113/0.125) of the variance in earnings between establishments in 1992 was due to firm fixed effects and 93.3%, (=0.140/0.150) of the establishment variance in 2007 was due to the firm fixed effects. Over time the variance in establishment earnings among establishments within firms fell, so that 47 percent of the increased earnings dispersion in the sample comes from increased earnings variance between firms.

Because many small firms have only a single establishment, however, our calculation that assigns all of the variance of single establishment firms to the firm could arguably overstate the dominance of firms in establishment effects. To see how much single-unit firms affect our finding that firm effects capture most of the employer impact on earnings we decomposed the variances of earnings among establishments for the multi-unit establishments in manufacturing over the 1992-2007 period and compare that decomposition to the decomposition for all establishments. Appendix table B gives the results of this exercise. It shows that among multi-establishment firms 83% (= 0.094/0.113) of the variation in establishment fixed effects is associated with the firm fixed effect compared to 89% (= 0.106/0.119) of the establishment variation associated with the firm fixed effect in a full sample. The bias is modest. Thus, our results are consistent with the emphasis of Song et al. (2015) on the importance of the *firm* in accounting for the increased dispersion in worker wages over time.

---

as many observations as possible from each individual for the panel data analysis where we identify both person effects and establishment effect, as the weakness of such identification is having few observations per individual.

10 This is a match to the so-called tfp-files (see Foster et al. 2016) as well as the LBD-files.

Cross Section Earnings Equations

Table 3 records estimated coefficients and standard errors for OLS regressions of the standard cross section earnings equation and of variants of our augmented earnings equation. Column (1) gives the coefficients and standard errors for the estimated effects on ln earnings of years of schooling, age, gender, and some interactions to allow for differences in effects among those attributes. In addition, the regression includes 171 geographic area dummies and 16 year dummies so that the coefficients are estimated within year and area. The estimated coefficients are similar to those typically found in the human capital earnings literature: an estimated average returns to years of schooling of about 9.4 percent per year and a concave age profile captured by the negative squared term and estimate gender and race wage gaps at 30% and 17%, respectively. The R^2 of the equation of 0.45 is larger than the R^2 in earnings functions fit on CPS data,[11] presumably because variation in earnings in the entire economy exceeds that in manufacturing and/or because the administrative LEHD earnings has less measurement error than self-reported earnings in the CPS.

Column 2 adds a set of workplace variables to reflect place of employment: 4-digit NAICS industry dummies, the ln number of employees of the *firm* and the ln number of employees in the establishment and establishment age and its square[12]. The estimates show significant firm- and establishment effects, and a concave wage-establishment age profile. Adding the firm and establishment characteristics raises the R^2 to 0.505 and thus explains 10% of the residual variance of earnings for demographically similar persons. The firm and establishment variables shrink the positive coefficients on years of schooling and age and the negative coefficients on gender and nonwhite, indicating that some of the impact of those factors comes through sorting of workers among establishments and industries within manufacturing.

Column 3 adds variables relating to the attributes of the establishment's work force: mean years of schooling, mean age, share female and share non-white; capital structures per worker and capital equipment per worker; the export share of establishment revenues; and the R&D investment of the firm to which the establishment belongs. The striking result is the high estimated coefficient on the years of schooling of all workers. The estimated nearly 7% increase in an individual's earnings for every year of average schooling of co-workers, above and beyond the 7.4 percent higher earnings boost from an extra year of the workers' own education suggests that it is almost as good to work in an establishment with more educated workers as it is to have more education. The estimates also show that workers earn more in establishments with older workers and less in establishments with a larger proportion of female or nonwhite workers; and that greater capital equipment per worker raises earnings more than greater capital structures per worker (a coefficient difference of 0.046 vs 0.007) and that earnings are higher in establishments with a high export share. Finally, earnings rise with R&D intensity of a firm: workers in firms with one standard deviation higher R&D intensity average 2% more earnings.

---

11 Estimating a similar regression with CPS data for the whole work force gave an R^2 of 0.35.

12 Dickens and Katz (1986) estimate industry wage differentials. Brown and Medoff (1989) study employer size-wage effect.

Column 4 gives the regression results with dummy variables for firms added to the equation while Column 5 gives results with dummy variables for establishments replacing those for firms. Addition of the firm fixed effects substantially reduces the estimated impact of the number of employees at the firm, indicating that short run changes in firm employment have little effect on wages, but only reduce the coefficient on number of employees at the establishment modestly. The Column 5 estimates with dummy variables for establishment also markedly reduce the coefficient for firm employment but leave a substantial effect of establishment employment on wages. With establishment fixed effects in the equation, the positive effect of establishment employment suggests that an establishment operates along a rising supply curve of labor for short term increases in employment.

Addition of the firm and establishment dummies naturally shrinks the estimated effect of firm and establishment variables on earnings. The Column 4 firm fixed effects regression eliminates the negative relationship between the share of non-white employees. Working in an establishment with a large non-white share is associated with low wages, but short run changes in the non-white share do not affect establishment earnings much. The Column 4 fixed effects regression also greatly weakens the relationship between R&D and earnings, reducing the estimated coefficient by over 80 percent. While R&D firms pay more than firms that do less R&D, changes in R&D activity within a firm has little impact on earnings.

The Column 5 regressions which include establishment fixed effects further shrink the coefficients of most of the establishment workforce characteristics compared to those in column 3. The column 3 estimated 0.0690 effect of the mean years of schooling on earnings drops to 0.0249 in column 5 while the estimated -0.3176 for being female in column 3 drops to -0.1084 in column 5. While measurement error usually accounts for some of the lower coefficient on variables in longitudinal analysis compared to cross section analysis (Freeman 1984), the pooling of observations to create average characteristics is likely to diminish measurement error so that huge drops in the effects of these characteristics are more likely to reflect economic behavior, as firms adjust earnings to changing characteristics gradually over time.

Panel Earnings Equations

The longitudinal structure of the LEHD allows us to estimate the effects of employer characteristics on earnings for the *same* individual in two ways: by comparing workers who remain in the same job over time while management changes characteristics of the establishment or accepts changes from other sources;[13] or by comparing workers who quit an employer with one set of characteristics to join an employer with other characteristics. Outside of recession years, the bulk of the labor mobility comes from worker decisions to move to a new employer willing to hire them. In recessions, mobility depends more on the decisions of firms, with the

13 Changes that management accepts refers to changes in attributes due to worker decisions, such as voluntary retirement where management did not seek to hire experienced workers as replacements or mobility where the management did not replace a worker who left with someone having the same skill or characteristic, and the like.

number of layoffs increasing to approach or exceed the number of quits[14]. While our data lack information on whether a worker left a job by quitting or by layoff, the fact that recession years are less frequent than non-recession years suggests that the bulk of the worker changes reflect quits rather than layoffs.[15]

Table 4 presents estimates of the effect of employer attributes on the earnings of the same worker when those attributes change. The first column shows the results of adding individual fixed effects to the basic ln earnings regression from column 3 in table 3. The coefficients on some employer level variables declines with the addition of the worker fixed effects: the estimated coefficient for average years of schooling of workers in an establishment falls by 59% (from 0.0737 to 0.0299), suggesting that much of the large co-worker schooling impact is due to positive sorting of workers by unmeasured individual characteristics into establishments with more educated workers. The coefficient on the equipment stock of capital per employee also drops massively by 70% (from 0.0462 to 0.0140), suggesting positive sorting of unmeasured individual characteristics into establishments with more equipment capital. And the coefficient on R&D drops by 83% (from 0.762 to 0.1290) , suggesting that most of the cross section R&D effect is due to a positive matching between R&D firms and unmeasured individual characteristics.

The next two columns unpack the fixed effects model into its two parts. Column 2 estimates the effect of employer characteristics on the wages of workers who stay in the same establishment. This controls for establishment fixed effects and match-specific fixed effects as well as for unobserved individual fixed effects. Column 3 estimates the impact of employer factors on the wages of workers who changed employers, which identifies the effects of establishment characteristics through changes in the employer, and thus does not control for establishment fixed effects or match-specific effects.[16]

For most establishment characteristics, the estimated effects of worker initiated changes in column 3 have a much greater impact on earnings than the estimated effects of employer-initiated changes in column 2. Moving to an establishment in a larger firm gives a wage increase of 0.0162 while a change in number of workers at a given establishment gives a 0.0056 boost to earnings -- about one-third as large. Moving from a less education intensive establishment to a more education intensive establishment increases wages by 0.0322 compared to an increase of 0.0048 in earnings for an increase in average education in a workers' current establishment. Moving to an establishment with older workers raises the wage of the mover while staying in an establishment with a rising age of the work force reduces the workers' wage. The effect of R&D

14 For non-recession years the number of quits divided by the number of layoffs exceeds 1.0 by 30%-50%. In recession years, the number of layoffs exceeds quits. http://www.bls.gov/jlt/jlt_labstatgraphs_oct2015.pdf, chart 7.

15 We did not probe possible differences between job changes from establishments having large drops in employment, where layoffs are potentially important, and job changes from establishments with stable or growing employment, where the locus would likely be voluntary shifts to better outside opportunities.

16 For this analysis we examine every job-to-job move in the data, retaining only the observations before and after the move, and estimate the regression including individual fixed effects.

on wages is more than twice as large for movers than for stayers (0.2075 vs 0.0860). But not all characteristics have a larger effect for movers than stayers. An increase in establishment employment has a modestly larger effect for persons who stay with an establishment than for those who move, and similarly for the share of non-whites.

Mechanically, the differences between the column 2 stayers-based estimates and the column 3 movers-based estimates reflect the fact that the stayers analysis controls for unobserved establishment fixed effects and thus removes correlations between those effects and the wages while the movers model does not do this. But the differences also reflect economic behavior. A worker who chooses to change employers will likely require a larger increase in pay than one who stays at a job as the work changes due to a changing workplace. An establishment that changes characteristics will likely adjust operations slowly and thus alter pay less in the short run than the pay difference between employers that have had different characteristics over longer periods.[17]

Earnings equations with individual fixed effects cannot identify the effect of fixed individual characteristics on earnings. But it is possible to learn something about how those characteristics affect earnings by regressing the estimated fixed effect for individuals on measures of individual characteristics. With say ten workers with two defining characteristics, say years of schooling and gender, the earnings equation would produce ten fixed effects to regress on schooling and gender.

Columns 1-3 of table 5 give the results of our analysis of the effect of fixed worker characteristics on the estimated fixed effects of workers in three fixed effect models with different structures. Model 1 estimated individual fixed effects without employer characteristics.[18] Model 2 estimated individual fixed effects from a regression with observable employer characteristics. Model 3 estimated individual fixed effects from a stayers' regression that includes establishment and match-specific fixed effects (eq. 3).

For individual effects positively related to the characteristics of employers, the estimates should decline across the columns, as they do. The returns to years of schooling drops from 0.1076 to 0.0841 from the model 1 specification to the model 3 specification with controls for observed and unobserved establishment effects. The coefficient on female falls by more than 10% and the coefficient on age falls by 18%.

The bottom line in Table 5 labeled "variance of the unobserved individual fixed effect", shows how the addition of establishment characteristics reduces the contribution of the fixed effects for individuals to the variation of earnings among workers. In model 1, the individual

17 Measurement error will also bias downward the estimates based on changes, for the basic reason that a given error will have proportionately larger impact on the small variation in year to year changes at the same workplace than on the larger differences between the employer the worker joins and the employer the worker leaves.

18 The difference is that in the fixed effects specification, the unobserved individual fixed effects are allowed to be correlated with all the included time-varying covariates.

fixed effect variance is 0.149, or 51% of the total variance. In model 2, which includes measured establishment characteristics, the individual fixed effect variance falls to 0.127 or 43% of the total variance. In model 3 with observed and unobserved establishment characteristics, the variance of the individual effect is 0.112, or about 38% of the total variance in earnings. Put differently, establishment factors account for 25% ((= 0.149- 0.112)/0.149) of the variance of estimated individual effects.

A full decomposition

Table 6 summarizes our findings with a full decomposition of ln earnings along the various dimensions in the augmented earnings equation. Standard individual characteristics (years of schooling, age, gender, and race) comprise 26% of the total variation in earnings; unobserved individual effects comprise 37% of the variation; observed establishment characteristics comprise 8%; unobserved establishment effects comprise 7%, and the match component 3%. The covariance between the individual and establishment components of the earnings equation adds 13% of the variance. The remaining variation arises from the transitory within-match residual comprising 7% of the total variation, in addition to the two small negative covariance terms between the observed an unobserved parts of the individual and establishment components respectively[19].

Given that the sorting of workers between establishments affects the returns to individual characteristics and earnings differentials by gender and race, we next analyze the ways in which workers and firms match up.

The sorting of workers between establishments

Table 7 gives the correlation coefficients that reflect sorting by key earnings determinants. The largest correlations show considerable homophily sorting of workers with similar characteristics: correlations of educated workers with educated workers (0.477), of older workers with older workers (0.333), females with females (0.349) and non-whites with non-whites (0.471). But other characteristics of employers are sufficiently correlated with worker characteristics to suggest sorting of workers among establishments beyond homophily. Educated workers work in large firms and in R&D intensive firms, in establishments with high capital per worker and with high export shares. These patterns make it likely that some of the education earnings premium comes through the greater likelihood that educated workers find jobs in employers with other wage enhancing characteristics. Older workers are also associated with establishments with high wage characteristics, though the correlations are much smaller. By

19 Our model is estimated under the assumptions that the fixed individual and establishment/firm effects remain constant throughout the sample period. Experiments with estimation on sub-periods show that this assumption is too restrictive, and that the variance of both the individual and establishment fixed effects rise over time during the sample period. The period over which to treat individual and establishment/firm fixed effects as fixed raises statistical and modeling issues that require analysis beyond the scope of the current study, on which we will report in a separate paper.

contrast, women work in establishments with lower capital intensity and non-white workers are largely employed in establishments with low wage characteristics.

The bottom two lines of the table shows the correlation between a measure of the establishment contribution to earnings through observed variables plus industry and region taken as a group, weighted by their estimated effect on earnings, and through establishment fixed effects with the individual characteristics. Both the establishment observables and fixed effects are most highly correlated with years of schooling, making schooling potentially the most important dimension of worker sorting among establishments.

Figure 1 summarizes the relations between the characteristics of workers and those of the establishments where they work in a different way. It displays the correlations between indices of the observed characteristics as a group, weighted by their respective coefficients in the earnings equation, and the fixed effects associated for workers as well as for establishments. The largest correlation is between the individual observables (weighted by their contribution to earnings) and establishment observables (0.253) (weighted by their contribution to earnings), followed by the correlation between the individual observables and the establishment unobserved fixed effect. By contrast, the fixed effect of individuals is weakly positively correlated with the establishment observables while the individual unobservables and unobserved establishment fixed effects are negatively correlated – a result consistent with Abowd et al. (2014). As Andrews et al. (2008) note, a negative correlation between two unobserved components of earnings could result from sampling and measurement errors,[20] so the safest conclusion from these correlations is that sorting of workers occurs largely on observable characteristics.

Mobility Among employers

The impact of employer characteristics on the earnings of workers with similar measured characteristics and fixed effects and the correlations in table 7 and figure 1 direct attention at the potential role of worker mobility among employers in determining pay. To what extent does mobility from job to job raise pay? Do workers who start their careers in establishments with low wage characteristics move to firms with better observable and unobservable characteristics over time? Conversely, how much downward firm mobility is there among workers who initially begin their careers at firms with high wage characteristics?

To examine the transitions of workers among establishments that differ in the establishment component of earnings, we formed a transition matrix for workers observed in our data both in 1992 and 2007. We attached to every worker in the sample the total establishment contribution of their employer to earnings, defined as the sum of the contribution to earnings of the time varying establishment characteristics, such as firm size and R&D spending, the fixed observables, such as industry and region, and the unobserved establishment effects. With an establishment contribution for each worker in 1992 and 2007, the natural measure of each workers' mobility is the change in the establishment component of earnings of their employer in those years.

20 Lise, Meghir and Robin (2013) give further discussion of these issues.

Table 8 summarizes the transition pattern by quintiles of the distributions, ordered from low-paying firms in quintile 1 to high-paying firms in quintile 5. The rows in the table show the distribution of workers by the quintiles of their establishment in the 1992 distribution of establishments into the quintiles of their 2007 employer in the distribution of establishments in that year. While the largest transition probabilities are for workers to remain in the same quintile over time, there is evidence of upward movement among establishments. Workers in the low quintiles have larger shares going up in the distribution than workers in the top quintiles have shares falling in the distribution. Among workers in the median group (3rd quintile) 38 percent move to a higher quintile, whereas only 21 percent move down while 40 percent remain in the same quintile. New workers come into the distribution of firms at the lower end and change jobs over time to produce a lifetime move up the distribution.

Finally, we characterize the sorting of workers with workers between establishments by Kremer and Maskin's (1996) index of segregation, $\rho = cov(\omega\ \underline{\omega})/V(\omega)$, where $\underline{\omega}$ is the average individual component of the establishment and $V(\omega)$ is the variance of the individual components of the standard earnings equation. If workers completely segregate between establishments according to their individual earnings components, $\rho = 1$. If they are randomly allocated between establishments $\rho = 0$. We calculated the index for both observable individual earnings components and for the unobserved fixed effect and obtained an index of 0.24 for observable characteristics, and 0.17 for unobservable characteristics. This supports the implication of the correlations that sorting of workers according to observed individual characteristics such as years of schooling, age, gender, and race is considerably stronger than segregation according to unobserved skills or other earnings attributes.

Characterizing the sorting of workers with establishments by the equivalent Kremer-Maskin (1996) index $\rho_{\Omega}=cov(\omega,\Omega)/V(\omega)$, where Ω refers to the earnings components of establishment, we divide the decomposition into its within (V^w) establishment and between (V^b) establishment parts by the identities:

(4a) $V^w(lnw) = V(\omega)(1-\rho) + V(\xi) + V(e)$

(4b) $V^b(lnw) = V(\omega)(\rho +2*\rho_{\Omega}) + V(\Omega)$

In our data, ρ=0.247 and ρ_{Ω}=0.100. The within establishment component is 59% of the variance in wages, of which 82% arises from the individual component (observed and unobserved), 6% from the match component, and 12% from the residual. The between component contributes 41% of the variance in wages, of which 36% is due to worker-worker sorting, 30% to worker-establishment sorting, and 34% to variance of the establishment effect.

Conclusion

This study has matched individual, establishment, and firm data-bases to estimate an earnings equation that augments standard regressions of ln earnings on the measured

characteristics of individual workers with measures of the characteristics of employers and with fixed effects estimates of unobserved characteristics of workers and employers. We find that:

1) Workers are paid more in establishments with more employees, in older establishments (up to a point), with greater equipment capital per worker and greater exports, with a workforce that has more educated workers, older workers, male and white workers; and in firms with greater R&D spending. Co-workers years of schooling has almost as large an impact on a workers earnings as the workers own years of schooling.

2) The estimated coefficients for employer characteristics diminish in longitudinal data when models include firm and establishment fixed effects, presumably because those models identify the effects from short term, possibly transitory, changes in characteristics while cross section differences reflect long term responses of earnings to characteristics.

3) Individual fixed effects models that identify coefficients based on the employer changing workplace characteristics for the same worker/job give markedly smaller though still generally significant estimates of the effect of characteristics on wages than fixed effect models that estimate coefficients from workers who change jobs.

4) There is considerable sorting of workers with similar workers, based on observable earnings characteristics and to a lesser degree on unobserved earnings characteristics and sorting of workers with establishments having similar high or low earnings attributes. The dynamics of worker mobility among employers has workers moving to enterprises with higher observable and fixed effects earnings components over time.

All told, bringing employers more into analysis of earnings illuminates the variance unexplained by the link between ln earnings and individual characteristics and illuminates the role of sorting in the relation between earnings and individual attributes. It also raises new questions for analysis: Why does having more educated co-workers have almost as large an effect on earnings as having additional education yourself? What mechanisms sort workers among firms in ways that increase education, gender, and race wage gaps? How much does the impetus for mobility – layoffs vs quits – affect worker outcomes? How much of firm effects on earnings can be linked to explicit wage policies? And given manufacturing's modest and declining share of employment, would augmented earnings functions in other industries tell us the same or different stories about the role of employers in determining earnings and in the sorting of workers among employers?

Table 1
Variance Decomposition of Ln earnings, All Sectors 1992 and 2007

	Variance Ln (earnings)		Share of variance	Change in variance	Share of growth
	1992	2007	Between establishments 2007	1992-2007	Between establishments 1992-2007
Manufacturing	**0.398**	**0.490**	**0.45**	**0.092**	**0.57**
Mining, Utilities, Transport	0.434	0.457	0.40	0.022	0.39
Business Services	0.612	0.713	0.56	0.101	0.86
Communication	0.502	0.634	0.40	0.132	0.53
Retail, Wholesale, Restaurants	0.508	0.551	0.48	0.044	0.80
Finance, Insurance, Real Estate	0.531	0.660	0.39	0.129	0.65
Private Services	0.427	0.482	0.49	0.054	0.90
Health, Education, Social Services	0.495	0.508	0.27	0.013	-0.15
ALL	**0.510**	**0.601**	**0.48**	**0.091**	**0.66**

Note: Numbers calculated from yearly regressions of log annualized sum of quarterly earnings for all jobs in the second quarter of the year on establishment dummies. Data from LEHD. Establishment is the sein unit.

Table 2
Variance Decomposition of Ln Earnings and in Manufacturing for Matched LEHD Panel, 1992 and 2007

	Variance 1992	Share	Variance 2007		Change 92-07	Share of change
Ln earnings	0.272	1	0.330	1	0.058	1.00
Between establishments	0.125	0.46	0.150	0.45	0.025	0.43
Between firms	0.113	0.42	0.140	0.42	0.027	0.47
Between estab. within firm	0.012	0.04	0.011	0.03	-0.001	-0.02
Within establishments	0.146	0.54	0.180	0.55	0.033	0.59

Note: Numbers calculated from a regression of log earnings on time dummies and establishment dummies. The matched sample include LEHD data matched to the Census of Manufacturers with valid observations of capital (from the ASM/CoM tfp-files, see Foster et al. 2016), to the education data from the Decennial Censuses and CPS, and that each individual is observed at least four times (see data description for details). All jobs included are observed in the second quarter of the year.

Table 3
Estimated Regression Coefficients and Standard Errors for Augmented Earnings Equations Including Firm and Establishment Characteristics for Manufacturing 1992-2007

	No estab. char.	Estab.char. I	Estab.char II	+ Firm fixed effects	+Estab. fixed effects
Years of schooling	0.0943***	0.0820***	0.0737***	0.0739***	0.0739***
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Age	0.0132***	0.0118***	0.0116***	0.0116***	0.0115***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Age square	-0.0005***	-0.0005***	-0.0005***	-0.0005***	-0.0005***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Female	-0.3022***	-0.2897***	-0.2691***	-0.2671***	0.2669***
	(0.0006)	(0.0006)	(0.0005)	(0.0005)	(0.0005)
Non-white	-0.1707***	-0.1539***	-0.1405***	-0.1397***	-0.1400***
	(0.0005)	(0.0005)	(0.0005)	(0.0005)	(0.0005)
Female x Age	-0.0050***	-0.0045***	-0.0045***	-0.0045***	-0.0044***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Female x Age^2	0.0000***	0.0000***	0.0000***	0.0000***	0.0000***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Female x Non-white	0.0309***	0.0318***	0.0384***	0.0371***	0.0373***
	(0.0009)	(0.0009)	(0.0008)	(0.0008)	(0.0000)
Establishment and firm characteristics					
Ln Firm employment		0.0295***	0.0173***	0.0071***	0.0012***
		(0.0001)	(0.0001)	(0.0005)	(0.0003)
Ln establishment employment		0.0266***	0.0270***	0.0258***	0.0232***
		(0.0002)	(0.0002)	(0.0003)	(0.0006)
Establishment age		0.0032***	0.0045***	0.0032***	-
		(0.0001)	(0.0001)	(0.0001)	
Establishment age squared		-0.0001***	-0.0001***	-0.0001***	-0.0000***
		(0.0000)	(0.0000)	(0.0000)	(0.0000)
Mean years of schooling			0.0690***	0.0483***	0.0249***
			(0.0003)	(0.0004)	(0.0007)
Mean age			0.0027***	0.0017***	-0.0034***
			(0.0001)	(0.0001)	(0.0001)
Share female			-0.3176***	-0.2548***	-0.1084***
			(0.0016)	(0.0029)	(0.0056)
Share non-white			-0.0185***	0.0029	0.0256***
			(0.0017)	(0.0028)	(0.0046)
Export share			0.0115***	0.0062***	0.0068***
			(0.0004)	(0.0005)	(0.0006)
Ln capital structures/employee			0.0074***	0.0033***	0.0018***
			(0.0002)	(0.0003)	(0.0004)
Ln capital equipment/employee			0.0462***	0.0211***	0.0075***
			(0.0003)	(0.0004)	(0.0005)
Firm R&D/employee			0.7260***	0.1228***	0.1627***
			-(0.0104)	(0.0131)	(0.0129)
Industry effects	-	Y	Y	Y	
Firm effects	-	-	-	Y	-
Establishment effects	-	-	-	-	Y
Region (171),year (16) effects	Y	Y	Y		
r2_adjusted	0.452	0.505	0.526	0.566	0.577
N	5.13E+06	5.13E+06	5.13E+06	5.13E+06	5.13E+06

Table 4
Firm and Establishment Characteristics. Individual Fixed Effects Models

	Individual FE	Job FE (spell)	Individual FE (Movers)
Ln firm employment	0.0161***	0.0056***	0.0162***
	(0.0001)	(0.0002)	-0.0003
Ln establishment employment	0.0305***	0.0214***	0.0180***
	(0.0002)	(0.0003)	-0.0005
Establishment age	0.0037***		-0.0008***
	(0.0001)		(0.0002)
Establishment age square	-0.0001***	-0.0000***	0.0000
	(0.0000)	(0.0000)	(0.0000)
Mean years of education	0.0299***	0.0048***	0.0322***
	(0.0003)	(0.0004)	-0.0008
Mean age	0.0023***	-0.0129***	0.0023***
	(0.0002)	(0.0001)	-0.0002
Share female	-0.1573***	-0.0036	-0.1765***
	(0.0020)	(0.0033)	-0.0044
Share non-white	0.0250***	0.1025***	0.0230***
	(0.0019)	(0.0029)	-0.0042
Export share	0.0000	-0.0021***	0.0020*
	(0.0003)	(0.0003)	-0.0009
Ln capital structures/employee	0.0059***	0.0002	0.0031***
	(0.0002)	(0.0003)	-0.0004
Ln capital equipment/employee	0.0140***	0.0027***	0.0189***
	(0.0002)	(0.0003)	-0.0006
Firm R&D expenses/employee	0.1290***	0.0860***	0.2075***
	(0.0060)	(0.0057)	-0.0188
r2_adjusted	0.873	0.913	0.827
N	5.13E+06	5.13E+06	7.31E+05

Note: All models include year dummies, age square, the interaction between gender and age, age square. The first and last models include individual fixed effects, and the second model includes job (unique combination of individual and establishment) fixed effects.

Table 5
Regression of Estimated Individual Fixed Effects on Fixed Individual Characteristics, from Three Models of Ln (Wage)

	Model 1: Fixed Effects from model with only individual characteristics	Model 2: Fixed Effects from model with individual & establishment characteristics	Model 3: Fixed Effects from model with individual & establishment characteristics and & establishment fixed effects
Years of schooling	0.1076***	0.0917***	0.0841***
	(0.0001)	(0.0001)	(0.0001)
Dummy variable for female	-0.3553***	-0.3326***	-0.3129***
	(0.0004)	(0.0004)	(0.0004)
Dummy variable for non-white	-0.1001***	-0.0882***	-0.1119***
	(0.0005)	(0.0005)	(0.0004)
Age	0.0136***	0.0128***	0.0112***
	(0.0000)	(0.0000)	(0.0000)
Interaction gender non-white	0.0504***	0.0500***	0.0429***
	(0.0009)	(0.0008)	(0.0007)
r2_adjusted	0.441	0.422	0.416
Variance of residual individual effects	**0.149**	**0.127**	**0.112**

Note: The dependent variable is the individual fixed effects from models including time varying covariates. Number of observations 5.13E+06 in all columns.

Table 6
Variance Decomposition of the Full Augmented Earnings Equation Model

Determinants of earnings	Variance decomposition		
Ln earnings	0.299		
Individual components	0.188		
Observed individual		0.078	
Unobserved individiual.		0.112	
2*Cov (Obs. indiv., Unobs.indiv.)		-0.002	
Within match residual	0.020		
Establishment components	0.043		
Observed estab.		0.024	
Unobserved estab.		0.020	
Unobserved firm			0.016
Unobs. estab. within firm			0.005
2*Cov (Obs. estab, Unobs.estab)		-0.001	
2*Cov(Indiv. comp., Estab.comp)	0.038		
2*Cov (Obs.ind., Obs.est.)		0.022	
2*Cov (Obs.ind., Unobs.est.)		0.012	
2*Cov (Unobs.ind., Obs.est.)		0.008	
2*Cov (Unobs.ind, Unobs.est)		-0.006	
Match component	0.010		

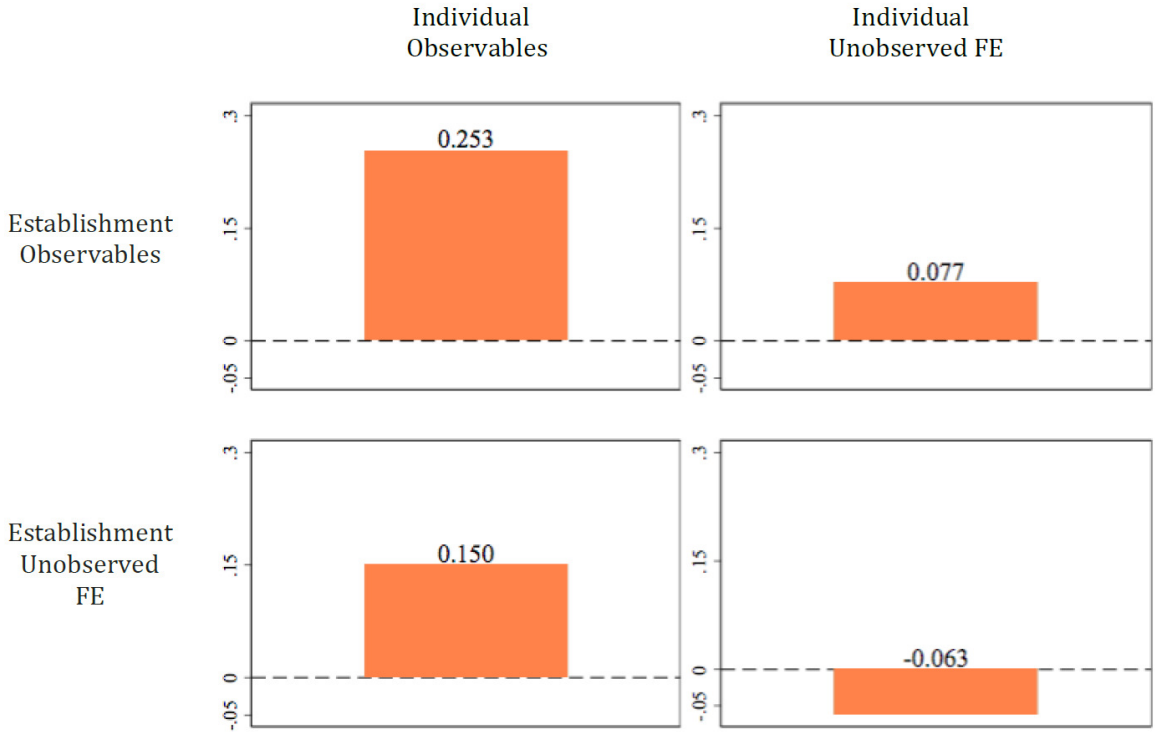
Note: Calculations using equation (3) to structure decomposition. Number of observations is 5.13E+06. Data for manufacturing matched sample, as described in text. Some number do not add up due to rounding errors.

Table 7
Correlation Coefficients between Individual and Establishment/Firm Characteristics

	Years of Schooling	Age	Female	Nonwhite
Ln firm employment	0.200	0.060	-0.002	-0.059
Ln establishment employment	-0.028	0.149	-0.014	-0.028
Establishment age	0.214	0.034	0.015	-0.042
Export share	0.103	0.044	0.005	-0.054
Ln structures capital/employee	0.185	0.069	-0.059	-0.087
Ln equipment capital/employee	0.117	0.071	-0.100	-0.075
Firm R&D/employee	0.216	0.015	0.000	0.000
Mean years of schooling	**0.477**	0.031	-0.030	-0.128
Mean age	0.110	**0.333**	-0.060	-0.036
Share female	-0.033	-0.041	**0.349**	0.105
Share non-white	-0.146	0.005	0.080	**0.471**
Establishment observables as a group, weighted by effect on earnings	0.258	-0.022	-0.084	0.030
Establishment fixed effect	0.112	0.069	-0.061	-0.066

Source: Tabulated from matched data file for manufacturing workers as described in text.

Figure 1: Correlation Coefficients between Individual Obervables and Unobservable with Establishment Observables and Unobservables



Source: Calculated from matched data file for manufacturing workers as described in text.

Table 8
Transitions of Workers Among Establishments Ordered by Establishment Contribution to Earnings (Observable Characteristics Weighted by Their Earnings Coefficients Plus Fixed Effect, by Quintile of the Distribution 1992-2007

2007 / 1992	Quintile 1	Quintile 2	Quintile 3	Quintile 4	Quintile 5	All	Change in Share	
							Up	Down
Quintile 1	0.564	0.258	0.097	0.056	0.024	1.000	0.436	
Quintile 2	0.172	0.401	0.287	0.108	0.032	1.000	0.427	0.172
Quintile 3	0.078	0.136	0.403	0.329	0.054	1.000	0.383	0.214
Quintile 4	0.042	0.062	0.113	0.451	0.331	1.000	0.331	0.217
Quintile 5	0.017	0.024	0.033	0.136	0.790	1.000		0.210

Note: Calculated on the balanced panel only. Quintiles of the distribution of the establishment effect include both unobserved and observed components of the establishment contribution.

Appendix Table A
Summary Statistics. Mean and Standard Deviation of Variables for the Full Sample of Persons in the LEHD Manufacturing Data Set and in the Matched Sample that Includes Measures of Years of Schooling

	Full Sample		Matched Sample	
	Mean	Std. dev	Mean	Std. dev.
Years of schooling after high school			0.723	2.311
Age	42.58	10.183	43.41	9.978
Female	0.300	0.458	0.294	0.456
Non-white	0.302	0.459	0.237	0.425
Ln firm employment	8.234	2.403	8.288	2.268
Ln establishment employment	6.300	1.572	6.317	1.512
Ln capital structures/employee	3.250	1.333	3.297	1.296
Ln capital equipment/employee	3.918	1.049	3.953	1.035
Export share of establishment	0.626	0.484	0.638	0.481
Firm R&D/employee	0.009	0.021	0.009	0.020
Ln earnings	6.649	0.554	6.682	0.543
Observations	23.4 million		5.1 million	

Note: Full sample tabulated for all workers in LEHD; matched sample for workers reporting years of schooling in match with Census or CPS as described in the text. Years of schooling measured as years after high school.

Appendix Table B
Variance Decomposition of Earnings Among Workers, in All Firms and in Multi-Establishment Firms, Matched Panel, 1992-2007

	All	Share	MU's	Share
Ln earnings	0.293	1	0.279	1
Between establishments	0.119	0.41	0.113	0.41
Between firms	0.106	0.36	0.094	0.34
Between establishments		0.04		0.06
Within firm	0.013		0.018	
Within Establishments	0.174	0.59	0.166	0.59

Note: Firms in the "All" establishment sample include the establishment effects for single-unit firms whereas the "MU's" sample only include multi establishment firms (defined as multi-unit firms within manufacturing only). Numbers calculated from a regression of log earnings on time dummies and establishment dummies. The total variance is calculated after subtracting variance due to the time dummies. Firm effects estimated from regression of establishment fixed effects on firm dummies. Multi-unit firms are defined as multi-unit firms within manufacturing only.

References:

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

Abowd, John M., Robert H. Creecy, and Francis Kramarz, 2002. Computing person and firm effects using linked longitudinal employer-employee Data, Longitudinal Employer-Household Dynamics Technical Papers 2002-06, Center for Economic Studies, U.S. Census Bureau.

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

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. Working Paper no. 13043, National Bureau of Economic Research, Cambridge, MA.

Andrews, J., L. Gill, T. Schank and R. Upward. 2008. High wage workers and low wages: negative assortative matching or limited mobility bias? *Journal of the Royal Statistical Society* Series A, 171, no. 3: 673-697.

Barth, Erling, Alex Bryson, James C. Davis, and Richard B. Freeman. 2016. It's where you work: increases in earnings dispersion across establishments and individuals in the U.S. *Journal of Labor Economics*, Special Issue dedicated to Edward Lazear 34:S2 (Part 2) (April):S67-S97.

Brown, Charles, and James Medoff. 1989. The employer size-wage effect. *Journal of Political Economy* 97 no. 5(5):1027-1059.

Card, David, Francesco Devicienti, and Agata Maida. 2014. Rent-sharing, holdup, and wages: evidence from matched panel data. *Review of Economic Studies* 81, no. 1:84-111.

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

Davis, Steve J. and John Haltiwanger. 1991. “Wage dispersion between and within U.S. manufacturing plants, 1963-1986. In Martin Baily and Clifford Winston, eds., *Brookings papers on Economic Activity, Microeconomics* (September):115-200, Washington DC: The Brookings Institution.

Devroye, Daniel and Richard B. Freeman. 2001. Does inequality in skills explain inequality in earnings across advanced countries? Working Paper no. 8140, National Bureau of Economic Research, Cambridge, MA.

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, UK: Basil Blackwell.

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

Dunne, Timothy, Lucia Foster, John Haltiwanger, and Kenneth R. Troske. 2004. Wage and productivity dispersion in U.S. 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. Working Paper no. 13351, National Bureau of Economic Research, Cambridge, MA.

Foster, Lucia, Cheryl Grim, and John Haltiwanger. 2016. Reallocation in the Great Recession: cleansing or not? *Journal of Labor Economics* 34 (S1): S293 - S331.

Freeman, Richard Barry. 1984. Longitudinal analyses of the effects of trade unions. *Journal of Labor Economics* 2, no. 1:1-26.

Groschen, Erica. 1991. Sources of intra-industry wage dispersion: how much do employers matter? *Quarterly Journal of Economics* 106, no. 3:869-884.

Gruetter, Max and Rafael Lalive. 2009. "The importance of firms in wage determination" *Labour Economics* 16, no. 2: 149-160.

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. 3:409-446.

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

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 (editors). 2009. *The structure of wages: an international comparison*. Chicago, IL: The University of Chicago Press.

Lise, Jeremy, Costas Meghir, and Jean-Marc Robin. 2013. Mismatch, sorting and wage dynamics. Working Papers no. 18719, National Bureau of Economic Research, Cambridge, MA.

Margolis, David N., and Kjell G. Salvanes. 2001. Do firms really share rents with their workers? IZA Discussion Paper no. 330, 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.

Mendes, R., G. van den Berg, and M. Lindeboom. 2010. An empirical assessment of assortative matching in the labor market. *Labour Economics* 17, no. 6:919-929.

Mincer, Jacob. 1974. *Schooling, experience and earnings*. New York, NY: National Bureau of Economic Research.

Moene, Karl Ove, and Michael Wallerstein. 1997. Pay inequality. *Journal of Labor Economics* 15, no. 3:403-30.

Mortensen, Dale T., Bent Jesper Christensen and Jesper Bagger. 2010."Wage and Productivity Dispersion: Labor Quality or Rent Sharing?" 2010 Meeting Papers 758, Society for Economic Dynamics.

Mulrow, Edward, Ali Mushtaq, Santanu Pramanik, and Angela Fontes, 2011. Assessment of the U.S. Census Bureau's Person Identification Validation System. NORC at the University of Chicago.[21]

Nordström 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, IL: University of Chicago Press.

Rastogi, Sonya and Amy O'Hara. 2012. 2010 Census match study. U.S. Census Bureau Report.[22]

Song, Jae, David J. Price, Fatih Guvenen, Nicholas Bloom, and Till von Wachter. 2015. Firming up inequality. Working Paper no. 21199, National Bureau of Economic Research, Cambridge, MA (Revised June 2015).

21 http://www.norc.org/PDFs/May%202011%20Personal%20Validation%20and%20Entity%20Resolution%20Conference/PVS%20Assessment%20Report%20FINAL%20JULY%202011.pdf

22 https://www.census.gov/2010census/pdf/2010_Census_Match_Study_Report.pdf