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

Abstract

We augment standard human capital earnings equations with variables reflecting attributes of the establishment and firm that employs a worker and find that the education of co-workers, capital equipment per worker, the industry in which the establishment operates, and the R&D intensity of firms impact worker earnings, among others. In longitudinal data, we find that workplace fixed effects contributes substantially to the overall dispersion of ln earnings, though by less than individual fixed effects. The estimated relations between observed and unobserved measures of the workplace account for much of the dispersion of ln earnings, which suggests that research on earnings should go beyond standard human capital determinants of pay to "bring the workplace back in" to the earnings equation.

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

First draft for discussion only
November 8, 2015

Standard earnings equations relate ln earnings to the human capital and demographic attributes of individuals while paying relatively little attention to the attributes of the employing firm or establishment that affect earnings. Per the title, this paper augments standard earnings equations with measures of the characteristics of the establishment and firm which employ a worker. It finds that a substantial proportion of the variation of earnings is associated with the establishment and firm so that persons with similar observed or unobserved skill are paid quite differently in the labor market. It identifies some observable characteristics of establishments and firms that underlie these effects.

The analysis combines a diverse body of data on establishments from the Census of Manufacturing with data on employees (using establishment identifiers linked to the Longitudinal Employer Household Dynamics (LEHD)) and data on firms (using firm identifiers on establishment files). We decompose the variance of ln earnings in longitudinal data into the component due to individual fixed effects and the component due to job/workplace fixed effects and estimate the relation between some observable attributes of establishments and firms on earnings and the relations among individual and establishment fixed effects.

Our decomposition shows that workplace fixed effects contribute substantially to the overall dispersion of ln earnings, though the impact is smaller than that of individual fixed effects[1]. We further find that some observable attributes of establishments – the education of co-workers and capital equipment per worker and the industry in which the establishment operates – and the R&D intensity of the firm that owns the establishment are associated with higher worker earnings and dispersion. These are attributes of the establishments and firms that are likely to be associated with the productivity of the workplaces, which again may affect the employers' ability to pay higher wages[2]. Our results hold under the two possible ways in which the association between workers and characteristics changes: when workers move to establishments/firms with larger amounts of specified attributes or when workers remain at the same establishment and the establishment/firm changes the attributes.

Overall, the evidence shows a considerable divergence of earnings from "the law of one price" in which the labor market gives each individual the same pay regardless of where he/she works[3]. Whether it is through rent-sharing of employer returns or matching of workers with employers, who you work for can greatly affect the dispersion of earnings.

Methodology

The traditional cross section human capital wage equation (Mincer 1974) relates earnings to observable measures/indicators of personal skill and other individual characteristics that can reflect productivity and/or employer attitudes or perceptions of the individual separate from productivity per se – ie prejudicial or statistical discrimination:

---

1 Early works that include unobserved workplace fixed effects are Groshen (1991), and workplace and individual effects Abowd et al (1999).

2 See eg Abowd (1989) Card et al (2010) Dobbelaere and Mairesse (2020), Margolis and Salvanes (2001), Martins (2009) Dunne et al (2004), Fagguioi et al (2007) Hellerstein et al (199) Lane et al (2007) for earlier works that link wage dispersion to the productivity dispersion.

3 It is possible to rescue the law of one price by assuming that any difference in wages over and above individual characteristics, are due to compensating differentials.

(1) $lnw_{ijt} = \beta_0 + x_{ijt}\beta + \gamma_t + u_{ijt}$

where x includes years of schooling and investment in skills, experience and individual attributes such as gender and race. Our augmented earnings equation adds establishment and firm characteristics to the equation:

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

where ψ_{ij} is a job/workplace fixed effect, defined as a unique combination of an individual and workplace.

Multiple observations on a single person and a single workplace in longitudinal data allow us to decompose the job/workplace effect into an individual fixed effect in terms of the estimated coefficient on a dummy variable for an individual, an establishment fixed effect, and a match component that we treat as orthogonal to the individual and establishment fixed effects, $\psi_{ij} = \alpha_i + \phi_j + \xi_{ij}$. Furthermore, we define the following two relations: $\alpha_i = X_i B + \alpha_i$, and $\phi_j = Z_j D + \varphi_j$, where X and Z denote covariates fixed for each individual and each establishment respectively.

In this structure, there is no way to identify B and D without assuming that the residual of the individual fixed effect is orthogonal to the individual fixed characteristics, and that the residual of the establishment fixed effect is orthogonal to the establishment fixed characteristics; although the components of the individual and establishment fixed effects can co-vary with the time varying characteristics as well with each other.

From the perspective of a "pure human capital model" in which 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. Alternatively, the coefficients of (1) can be viewed as biased estimates of the net effects of skills/attributes in equation (2) due to correlations between the covariates in equation (1) and the additional equation (2) variables.

The purpose of our analysis is to provide estimates of the establishment and firm specific variables in equation (2), and to see how accounting for these effects changes the estimates of the other coefficients and to provide estimates of the contribution of the increased dispersion in ln earnings among establishments to the increased inequality of earnings among workers in the past several decades (Barth, Bryson, Davis, and Freeman (2014)) that take account of the longitudinal data across Censuses of Manufacturing.

Census-LEHD & BRB Matched Data

To augment cross section and longitudinal ln earnings functions we combine several data files for workers, establishments, and firms. We limit our analysis to manufacturing because Census and other files provide more information for that sector than for others, but the calculations can be generalized to other sectors such as construction or services using the available data from Censuses of Production and other data on the attributes of establishments and firms.

The key dependent variable in our analysis are earnings for individual workers from the Longitudinal Employer-Household Dynamics (LEHD) Employment History Files for the nine states which have LEHD data from the early 1990's through 2007.[4] The LEHD data are linked to the quinquennial Census of Manufacturers (CoM) for the four economic census years, 1992, 1997, 2002 and 2007 using the LEHD Business Register

4 These states cover xx% of employment. Comparisons with data for states that cover different time periods show that the nine state sample are reasonably representative

Bridge (BRB). The BRB is a one to one link at the firm level. This creates a potential problem linking establishments in the CoM and in the LEHD because LEHD unemployment insurance reporting units are linked by firm, detailed industry and county to CoM establishments. For most observations, mapping from LEHD establishment to CoM establishment is unique so there is no problem. This includes cases of single-unit firms, firms operating at only one location, and plants that are not co-located with other same industry plants within the same firm. But for plants co-located with others in the same industry and firm, the link is not one-to-one. For these establishments, we aggregate total and average 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 data by linking workers to their reported characteristics from 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 same person identifier in the Census and CPS surveys as in the LEHD data. Recent decennial files, including the 2000 data, have very high PIK rates of 90-93% (Mulrow, Mushtaq, Pramanik and Fontes 2011, Rastogi and O'Hara 2012). However the 1990 PIK is more limited due to the available vintage of the address files.[5] 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.[6] We converted occupation codes to five job classes: science and engineering workers, managers, professional, clerk and administrative, and production workers.

The quinquennial Census of Manufacturing provides production-related data on all establishments in that sector, which we add to the files on employees. These data give measures of the number of workers at establishments, the proportion of workers who are production workers, establishment expenditures on labor, measures of establishment-level capital equipment and building stock, plant size and age, etc. We averaged the years of schooling of workers in an establishment to create an establishment-level years of schooling that represents its human capital intensity.

To obtain measures of the characteristics of the firm owning an establishment we make use of the Business R&D and Innovation Survey (BRDIS) on whether or not a firm reports R&D expenditures and the intensity of its R&D activity. For publicly owned/listed firms additional data is available from diverse sources that can provide insight into corporate practices, policies, and expenditures that may affect earnings.

Basic Variance Decomposition

As a first step in assessing the importance of workplaces on earnings, we decomposed the variance of the earnings of individuals among establishments and firms using a cross section design. We regressed ln earnings of individuals from the LEHD on time dummies and establishment dummies, where the establishment dummies are identified by having multiple individuals at every establishment. The part of the variance

---

5 Name and address files for individuals are highly sensitive, and are not generally distributed throughout the Census Bureau with the data files. The versions of 1990 decennial files used did not have the original name and address data, which had to be reconstructed with other available datasets before the file could be PIKed. As a result, the PIK matches favor less mobile adult heads of household.

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

associated with the establishment dummies is our estimate of the establishment effect. The part of the variance that is not accounted for by the establishment effect is, by definition, the within establishment effect. It is associated with differences in earnings for workers within an establishment. To determine the extent to which the establishment effects are attributable to establishments being part of the same firm, we regress the establishment fixed effects on the dummies for firms. The proportion of the variance in establishment fixed effects attributed to the firms reflects the effect of firms while the remaining proportion is the within firm establishment effect.

There are limits to such a decomposition analysis. First, many generally smaller firms have only a single establishment which makes the firm and establishment effects one and the same. To deal with this problem we differentiate multi-establishment firms from single establishment firms and decompose the variances separately for the two groups.[7] Second, the data include establishments and firms that appear in different numbers of CoM years. Some establishments/firms may be in the sample once while others establishments/firms may appear in two, three, or all four of the Census of Manufacturers. While we allow for covariance between establishment fixed effects and time, the panel is unbalanced and thus complicates our decomposition models. A third limit to our analysis is that our decomposition does not identify standard individual fixed effects from a regression of ln earnings on dummy variables for individuals (for whom we have at least two or more earnings observations). The within-establishment variance includes workers who appear in the data only once because they worked for a manufacturing establishment in only one of the CoM years. They may have worked in some other sector or been without work covered by unemployment insurance in other years. The within-establishment variance thus reflects all factors that cause differences in earnings among individuals in any given CoM year – potential fixed factors associated with measured and unmeasured skill of workers, variation in earnings of individuals around their fixed factor, and the unbalanced sample design. To see the extent to which these factors influence our decomposition, we run the decomposition on the full data set and on a data set that includes only firms and workers who appear in all four of the CoM years under study.

Table 1 gives our decomposition results. Panel A gives the results for the full data set in which we had the LEHD earnings for workers for manufacturing establishments. The columns labeled all establishments give the decomposition for all observations, including single unit firms, in the four CoM years. The decomposition shows that 58.2% (= 0.139/0.239) of the variance of log earnings is associated with earnings variation

7 The following shows the distribution of number of firms and workers by number of plants in a firm:

plants in a firm	firms	workers
1	31400	743000
2	1800	187000
3	600	111000
4	300	90000
5	200	63000
6+	600	714000

across workers within establishments whereas the remaining 42 percent is due to variation in earnings across establishments. The second column under all establishments shows that 89% (= 0.119/0.133) of the establishment variation is due to variation across firms, where the variation among single unit firms is counted as due to the firm rather than the establishment. Since in single unit firms, there is no variation across establishments, the columns labeled multi-unit establishments give the same decomposition for the firms where the CoM provides more than one establishment. Here, 84% (= 0.078/0.093) of the variation across establishments is associated with the firm fixed effect. The implication is that firm wage policies are a huge factor in earnings differences among establishments, while establishments within the same firm showing relatively modest variation.

The calculations in Panel B for the balanced panel by establishment and individual give similar results: a large variation in earnings associated with establishments, with the bulk of the establishment variation associated with the firm that owns the establishments, but with larger variation in earnings among persons working within establishments.

Cross Section Earnings Equations

Table 2 records estimated coefficients and standard errors for OLS regressions of the standard cross section ln earnings equation. The right hand side variables in column 1 relate to the attributes of an individual --- years of completed education, "experience" measured as age-years of schooling -five, experience squared, to capture the concave relation between age/years in the job market and earnings, dummy variables for gender, and for being nonwhite, and interactions between gender and the experience profile and being nonwhite. The results comparable to those typically found in the human capital earnings literature: an estimated average returns to education of 10 percent per year, a concave experience profile captured by the negative squared experience term, and gender and race wage gaps. The R-square of the equation is .44[8]. Column 2 adds four digit NAIC industry dummies and the ln of the number of employees of the firm and the ln of the number of employees in the establishment[9]. This raises the R-square to 0.50 and shrinks the positive coefficients on years of schooling and experience and the negative coefficients on gender and nonwhite, indicating that some of the impact of those factors works through the channel of size of firm and establishment and within-manufacturing industry mix.

Column 3 adds further establishment characteristics and the R&D intensity of the firm to which the establishment belongs. The establishment characteristics relating to the work force in the establishment are its average years of schooling, experience, percentage female and percentage non-white. The estimated coefficients show that individuals who work in establishments which have a more educated work force earn more regardless of their own years of schooling, by about 5 percent for every year of schooling above and beyond the 8.2 percent higher earnings from an extra year of their own education. By

8 Estimating a similar regression with CPS data for the whole US labor market, we obtain an R square of . 35. The higher R-square in table 2 may reflect the smaller variation of earnings in manufacturing than in the entire CPS and/or smaller measurement error in administrative LEHD earnings than in self-reported earnings in the CPS.

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

contrast, workers in establishments with higher average experience are paid less than those in establishments with lower average experience, while earnings are also lower in establishments with a larger proportion of woman or nonwhites workers.

The establishment capital intensity is measured by the value of capital structures per employee and the value of equipment per employee. The effect of higher capital per worker on ln earnings occurs largely through greater capital equipment per worker with an estimated elasticity of 0.034 rather than through greater capital structures per worker with an elasticity of just 0.002 for capital structures and 0.034 for capital equipment.

The R&D intensity of the firm which owns the establishment is measured by dummy variables relating to R&D employment per worker. The dummy variables are for the quintile in the distribution of firms which do R&D. The deleted group are firms with no R&D. Workers in firms that do R&D are paid more and earnings rise with the R&D intensity.

In sum, observable measures of the characteristics of establishments and firms contribute significantly to explaining the variation in earnings among individuals, though they contribute less than do the characteristics of individuals.

The longitudinal nature of the LEHD allows us to estimate the effects of establishment factors on earnings for the same individual over time and thus to test the possibility that the relation between establishment factors results from selectivity of persons into establishments on unobserved attributes. For workers who remain in the same job over time, we can compare earnings in establishments that changed the size or characteristics of the work force, or whose firm changed its size and R&D intensity over time. For workers who changed jobs over time, we can compare earnings in the establishment in which the worker initially worked with that in the establishment to which the workers moved.

The calculations in Table 3 present the results of such fixed worker analysis. The first column shows the results of adding individual fixed effects to the regression of ln earnings on the attributes of an establishment and firm. The coefficients are identified by the variation in characteristics for each individual over time rather than from cross section comparisons of different workers at workplaces with differing establishment/firm attributes. The coefficients from the fixed effects analysis differ from those estimated from the cross section – with those on establishment education and firm R&D in particular being somewhat smaller – but still have a clear impacts on individual earnings. The implication is that the establishment/firm effects are largely "real" as opposed to artifacts due to selectivity of persons with unobserved skills into workplaces with those attributes.

In the context of our decomposition analysis, however, the overall fixed effects result in column 1 are imperfect measures of the role of individual unobserved characteristics, as they could be affected by unobserved establishment characteristics as well as the unobserved individual effects. Accordingly, we differentiate in columns 2 and 3 the impact of the establishment/firm factors depending on whether an establishment/

firm attribute changed for workers staying the same workplace (ie for a fixed employer/ firm) or for workers who moved to a new employer (ie across workplaces). Column 2 contrasts the earnings of the same worker in the same workplace, which captures unobserved individual fixed effects, establishment fixed effects and potential match specific fixed effects as well. It identifies the effect of establishment/firm attributes on wages solely through employer initiated changes in the attributes or to changes due to worker mobility that management did not obviate through offsetting hiring decisions.[10] Most of the column 2 coefficients are similar to those in column 1 but the estimated impact of establishment size increases in column 2 compared to column 1 while the impact of average years of schooling in the establishment drops to small and insignificant and the coefficient for the share of women flips sign.

In column 3 we contrast the earnings of the same worker when he or she moves across establishments with differing values of the establishment/firm characteristics. We pick up every job-to-job move in the data, and retain the observation before and after the move, and estimate the same regression, including individual fixed effects. This analysis identifies the effects of those characteristics on wages solely through changes in the workplace, presumptively largely initiated by the worker.[11] The results bridge the gap between the estimates in the first two columns. Moving to an establishment with a higher educated work force gives a large increase in wages, while moving to a female dominated establishment is associated with wage decline. On the other side, moving among establishments of a different size has a smaller impact on wages than being in a given establishment that experiences a large change in employment.

There are alternative possible reasons for the differences between the column 2 stayers-based estimates and the column 3 movers-based estimates based on their use of different dimensions of the data to identify establishment/ firm effects that further work can illuminate. One difference is the fact noted earlier that the stayers model controls for unobserved establishment fixed effects which the movers model does not do. This menas that the column 2 estimates remove any biases due to correlations between the covariates and any unobserved fixed establishment effect. Another possible cause of differences between two estimates is in differences in the timing of adjustments between an establishment which has changed a characteristic and establishments with long-standing differences in the characteristic. For instance, in the time period covered in our analysis a change in the share of high educated workers in a given establishment may raise wages less than a comparable change in the share of highly educated workers from a worker moving from an establishment with a smaller to share of educated workers to one with a larger share, because it takes time for the establishment that is upgrading its work force to adjust its mode of operating to the change. It is also likely that there are different measurement error problems: the variation in year to year changes in the mean years of schooling or R&D intensity is likely to be relatively more contaminated by error than the larger differences that occur when workers change establishments or firms.

---

10 By changes that management accepted, we mean such developments as a decline in a workplace attribute, say the experience of the work force due to retirement, where management did not seek to hire experienced workers as replacements.

11 We could but have not probed this point further by contrasting shifts associated with large drops in establishment employment, where the locus of the workers' shift would likely be a shrinkage of demand and shifts from establishments with stable or growing employment, where the locus would likely be workers following better outside opportunities.

Fixed individual effects and characteristics

By construction, earnings equations that exploit longitudinal changes to estimate individual fixed effects cannot identify the role of observed characteristics of individuals that are fixed over time on earnings. What is possible, however, is to relate estimated individual fixed effects to the fixed characteristics of the individual and thus to obtain the average relation between the fixed effects and characteristics. In table 4 we use estimates of individual effects from different regression models as dependent variables and estimate their relation to fixed individual attributes. The estimated individual effect in column 1 comes from a regression that excludes any establishment characteristics. It thus should mirror the results from the table 2 ordinary least squares regression of ln earnings on the individual attributes. The estimated individual effects in column 2 come from a model that includes individual and establishment/firm observables and should be smaller than those in column 1 to the extent that individual effects are related to the characteristics of establishment/firm. Finally, the estimates of individual effects in column 3 are from a model that includes a full set of individual and establishment fixed effects. Here the individual effects are obtained by first estimating the coefficients for the time varying covariates within job (model two in Table 3), then decomposing the job/workplace effect into a person and an establishment effect producing the following decomposition model:

$$(3)\ lnw_{ijt} = \beta_{0t} + x_{ijt}\beta_t + \gamma_t + z_{jt}d_t + X_iB + \alpha_i + Z_jD + \varphi_j + u_{ijt}$$

The estimated coefficients on individual fixed characteristics change as we take increasing account of the role of establishment/firm characteristics. The returns to years of schooling drops from .12 to .09 when we move from a specification without establishment characteristics to a model with full control of all observed and unobserved establishment effects. The coefficient on female drops by more than 10 % and the coefficient on years of experience falls by 18% while the coefficient on non-whites increases by 25%, primarily when the calculation takes account of unobserved establishment effects. This indicates some differential sorting of workers with differing individual fixed effects among establishments with differing fixed effects.

Establishment and firm effects over time

To what extent have establishment/firm fixed effects changed over time as earnings inequality has risen? Comparing cross section data over time using data from the LBD, Barth, Bryson, Davis & Freeman (2014) found that on the order of 2/3rds of the upward trend in dispersion of earnings is associated with rising establishment differentials. How well does this hold up in the context of models that estimate establishment and individual effects in the Census of Manufacturers data? Identifying both establishment and individual fixed effects is demanding in the CoM because we have at most four observations over the 1995 - 2007period under study.[12] Many establishments and individuals appear fewer than four times in the data and only a small number make it into a balanced individual and enterprise panel for the entire period. Identification of the fixed effects has too few observations per fixed effect to allow for a

12With yearly data, we could estimate fixed effects on, say the first and last 5 years, and then compare the variances of the fixed effects etc (see eg Card et al).

standard regression of earnings on dummy variables for establishments and individuals to produce reliable results[13]. In addition, attrition due to our decennial census match that could affect the estimates of the variance among the fixed effects over time.

However, we may utilize our data with the following simplification of the model with time varying effects. We assume that the fixed effects are constant over time, but have a potentially different impact on wages every point in time. Consider the following representation of equation (1)

$$(4)\ lnw_{ijt} = \beta_{0t} + x_{ijt}\beta_t + \gamma_t + z_{jt}d_t + X_iB + \tau_t a_i + Z_jD + \theta_t\varphi_j + u_{ijt}$$

where $u_{ijt} = \xi_{ij} + e_{ijt}$ is assumed to be orthogonal to the variables in the model. τ_t and θ_t allow the person and establishment fixed effects to have a different impact on earnings each period of time. Normalizing τ_t and θ_t to be unity on average over the sample period, our estimation strategy involves a three step procedure: Step 1, where we estimate the coefficients of time varying covariates within a worker-establishment match on the full sample, and next decompose these into fixed individual and establishment effects, a_i and ϕ_j, ala AKM (see Abowd et al and Abowd et al). Step 2, in which we regress the estimated fixed effects on individual and establishment measurable characteristics that are constant over time. The residual of the latter equations are our estimate of the unobserved fixed effects a_i and φ_j . Step 3: Estimate (4) separately each year, including each of the two estimated fixed effect from step 1 as a covariate in the equations. This provides us with time varying coefficients for all the observed variables, as well as an estimate of the impact factors τ_t and θ_t .

An assessment of the change in the variance of the establishment and person fixed effects between periods of time may then be provided by noting that the variance for the individual effect each year is given by $V(\tau_t a_i) = \tau_t^2 V(a_i)$, and similarly for the establishment fixed effect $V(\theta_t\varphi_j) = \theta\tau_t^2 V(\varphi_j)$, and that it is thus sufficient to consider the estimate for the impact factors without having to calculate the variance each year.

Table 5 provides the year-by-year results. Panel A gives results for the entire sample. The variances of the establishment effects and the individual effects increase by about 12 percent over the period. And the impact of the establishment and firm characteristics, reflected in the estimated coefficients on the covariates increase over time as well. (1.026^2/0.97^2) =1.12. Panel B gives the results for the sample of individuals who show up at least 4 times in our data. Within this subsample, the variance of both the unobserved fixed individual and establishment effects increase by more than 50 percent from 1992 to 2007. ((1.901^2)/(0.089^2)=1.5 These increases are sizable compared to the percent increases in standard measures of inequality, such as the variance of ln earnings for individuals.

13 Note that while the precision of the "within" estimators of the observable characteristics is determined by the number of observations, the precision of each fixed unobserved effect is determined by the number of times the unit is observed.

Conclusion

This study has matched individual, establishment, and firm data bases that allow us to link measured characteristics of individual workers, unmeasured characteristics inferred from their earnings history with measured characteristic of establishments/firms and unmeasured characteristics inferred from their earnings over time to assess the importance of measured and unobservable firm/establishment effects on individual earnings. Our decomposition analyses find that variation in individual fixed effects contribute more to the variance of earnings than the variation in establishment fixed effects, but that the variation in establishment fixed effects is a major factor in earnings; that much of the establishment fixed effects are associated with the firm that owns the establishments, which directs attention at firm wage and labor policies. Our analysis of of the impact of observable attributes of workers and establishments in earnings regressions yield qualitatively similar results, with substantial effects for establishment/firm characteristics that vary depending on the source of variation of changes in the characteristics. Finally, our analysis of changes of fixed establishment effects yields large estimates that show that increases in the dispersion of those effects are large enough to have played a big role in the rise of inequality.

Table 1. Variance decomposition of ln earnings, firms, establishments, and workers within establishments, Census of Manufacturers, 1992, 1997, 2002, 2007.

Panel A : All observations in CoM

	All Establishments		Multi-Unit Establishments Only	
Var ln earnings	0.335	0.335	0.303	0.303
Due to Establishments	0.133		0.122	
Due to Firms		0.119*		0.099
Due to establishments within firm		0.014		0.021
Within establishment	0.202	0.202	0.181	0.181
Number of observations	1,907,000	1,907,000	1,121,000	1121000

Panel B: Observations that provide a balanced establishment-worker panel

	All Establishments		Multi-Unit Establishments Only	
Var ln earnings	0.239	0.239	0.205	0.21
Due to Establishments	0.100		0.093	
Due to Firms		0.093*		0.078
Due to establishments within firm		0.008		0.014
Within establishment	0.139	0.139	0.112	0.112
Number of observations		225,000		127,000
individuals	225,000		127,000	

*Firms in the All establishment sample include the establishment effects for single-unit firms.

Note: 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-sunit firms within manufacturing only.

Table 2 Estimated Regression Coefficients and Standard Errors in parenthesis for OLS Estimates of ln earnings equations for manufacturing employees 1992, 1997, 2002, 2007.

	1	2	3
Years of Schooling	0.1037 (0.0001)***	0.0901 (0.0001)***	0.0824 (0.0001)***
"Experience" (Age-Yrs Schooling -5)	0.0198 (0.0001)***	0.0181 (0.0001)***	0.0181 (0.0001)***
"Experience" squared	-0.0006 (0.0000)***	-0.0006 (0.0000)***	-0.0006 (0.0000)***
Dummy variable for Female	-0.3363 (0.0010)***	-0.3239 (0.0009)***	-0.2937(0.0009)***
Dummy variable for Nonwhite	-0.1906 (0.0010)***	-0.1782 (0.0009)***	-0.1652 (0.0009)***
Interaction Female x "Experience"	-0.0069 (0.0001)***	-0.0063 (0.0001)***	-0.0063 (0.0001)***
Interaction Female x "Experience" sqrd	0.0001 (0.0000) ***	0.0002 (0.0000)***	0.0002 (0.0000)***
Interaction Female x Nonwhite	0.0534 (0.0015)***	0.0537 (0.0015)***	0.0612 (0.0014)***
Log number workers in firm		0.0206 (0.0002)***	0.0072 (0.0002)***
Log number of workers in establishment		0.0320 (0.0003)***	0.0363 (0.0003)***
Establishment Characteristics			
Mean yrs of schooling			0.0540 (0.0004)***
Mean "experience" of workers			-0.0016 (0.0001)***
Mean percent woman			-0.2789(0.0021)***
Mean Percent nonwhite			-0.0139 (0.0024)***
Ln Capital per employee Structures			0.0021 (0.0003)***
Ln Capital per employee Equipment			0.0335 (0.0004)***
Firm R&D/Employment Quintile With No RD as deleted group			
Bottom quintile			0.0125 (0.0011)***
2nd lowest quintile			0.0361 (0.0012)***
Middle quintile			0.0653 (0.0012)***
2nd highest quintile			0.0822 (0.0014)***
Top quintile			0.0717 (0.0016)***
Industry fixed effects	-	Y	Y
Regional fixed effects	Y	Y	Y
Year fixed effects	Y	Y	Y
r2_a	0.442	0.496	0.514
Number of observations	1.91e+06	1.91e+06	1.91e+06

Note: The models also include dummy variables for 171 PMSAs and local labor markets surrounding them within each state, and a time trend.

Table 3 Estimated coefficients and standard Errors of Ln Earnings on establishment and firm characteristics with fixed effects for individuals, for workers at the same establishment ("stayers") and workers who move among establishments ("movers")

	FE individual	FE job/workplace: Stayers	FE Movers
Log # workers in firm	0.0117***	0.0070***	0.0144***
	(0.0003)	(0.0005)	(0.0004)
Log # workers in establishment	0.0401***	0.0752***	0.0286***
	(0.0005)	(0.0010)	(0.0006)
Mean yrs of schooling	0.0116***	-0.0013	0.0253***
	(0.0006)	(0.0008)	(0.0009)
Mean "experience" of workers	-0.0031***	-0.0035***	-0.0017***
	(0.0001)	(0.0001)	(0.0002)
Mean percent woman	-0.0908***	0.0173***	-0.1536***
	(0.0029)	(0.0042)	(0.0044)
Mean percent nonwhite	0.0432***	0.0448***	0.0489***
	(0.0031)	(0.0045)	(0.0048)
Ln capital stock per worker	0.0020***	0.0012*	0.0031***
	(0.0003)	(0.0005)	(0.0005)
Ln equipment per worker	0.0103***	0.0092***	0.0097***
	(0.0004)	(0.0006)	(0.0007)
Bottom quintile RD/employment	0.0077***	0.0043***	0.0104***
	(0.0009)	(0.0010)	(0.0018)
2nd lowest quintile RD/Employment	0.0041***	-0.0018	0.0094***
	(0.0011)	(0.0012)	(0.0019)
Middle quintile	0.0120***	-0.0017	0.0250***
	(0.0012)	(0.0014)	(0.0021)
2nd highest quintile	0.0231***	0.0215***	0.0220***
	(0.0013)	(0.0016)	(0.0024)
Top quintile of RD/Employment	0.0174***	0.0264***	0.0139***
	(0.0016)	(0.0016)	(0.0031)
r2_a	0.881	0.904	0.827
N	1.91e+06	1.91e+06	4.26E+005

Note: The models also include all time varying variables in Table 2 with the exception of the linear term for experience, which is not separately identifiable from a linear projection of the time dummies.

Table 4 Estimated Coefficients and Standard Errors for Regressions of Individual fixed effects on individual characteristics in models with measured structures.

	Models With Different Measures Beyond Individual Measures		
	Only Individual characteristics	Individual & establishment characteristics	Individual & establishment characteristics and individual & establishment fixed effects
Years of schooling	0.1161***	0.1030***	0.0942***
	(0.0001)	(0.0001)	(0.0001)
Dummy variable for female	-0.4105***	-0.3874***	-0.3569***
	(0.0008)	(0.0008)	(0.0007)
Dummy variable for non-white	-0.1136***	-0.1133***	-0.1438***
	(0.0009)	(0.0009)	(0.0009)
Years of experience	0.0195***	0.0187***	0.0161***
	(0.0000)	(0.0000)	(0.0000)
Interaction gender non-white	0.0691***	0.0637***	0.0524***
	(0.0016)	(0.0015)	(0.0015)
r2_a	0.438	0.429	0.392
Variance of residual individual effects			0.167
N	1.91e+06	1.91e+06	1.91e+06

Table 5. Estimated Coefficients and Std Errors on Personal and Establishment Factors in Ln Earnings Equations Over time for models that include fixed individual and establishment effects

	Panel A: Full Sample for each year			
	1992	1997	2002	2007
Years of Education	0.0915***	0.0931***	0.0957***	0.0981***
	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Experience (Age-Yrs Schooling -5)	0.0168***	0.0165***	0.0154***	0.0150***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Experience squared	-0.0007***	-0.0006***	-0.0006***	-0.0006***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Dummy variable for Female	-0.3441***	-0.3478***	-0.3404***	-0.3465***
	(0.0004)	(0.0005)	(0.0006)	(0.0007)
Dummy variable for Nonwhite	-0.1307***	-0.1253***	-0.1239***	-0.1229***
	(0.0005)	(0.0005)	(0.0006)	(0.0006)
Interaction female with Experience	0.0014***	0.0019***	0.0029***	0.0032***
	(0.0000)	(0.0000)	(0.0001)	(0.0001)
Interaction female with Experience^2	0.0002***	0.0001***	0.0001***	0.0001***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Interaction female with Nonwhite	0.0116***	0.0032***	-0.0053***	-0.0106***
	(0.0007)	(0.0007)	(0.0009)	(0.0010)
Log number workers in firm	0.0073***	0.0072***	0.0088***	0.0055***
	(0.0001)	(0.0001)	(0.0001)	(0.0002)
Log of workers in establishment	0.0743***	0.0743***	0.0723***	0.0768***
	(0.0002)	(0.0002)	(0.0002)	(0.0002)
Mean yrs of schooling	0.0002	0.0013***	-0.0022***	-0.0038***
	(0.0002)	(0.0002)	(0.0003)	(0.0003)
Mean yrs of experience	-0.0034***	-0.0032***	-0.0036***	-0.0038***
	(0.0000)	(0.0000)	(0.0001)	(0.0001)
Mean percent female	0.0027**	0.0140***	0.0238***	0.0312***
	(0.0010)	(0.0011)	(0.0014)	(0.0016)
Mean percent nonwhite	0.0278***	0.0404***	0.0464***	0.0598***
	(0.0012)	(0.0013)	(0.0016)	(0.0017)
Ln capital structure per worker	-0.0002	0.0013***	0.0005**	0.0013***
	(0.0002)	(0.0002)	(0.0002)	(0.0002)
Ln equipment per worker	0.0111***	0.0089***	0.0087***	0.0094***
	(0.0002)	(0.0002)	(0.0003)	(0.0003)
Lowest RD/employment quintile	0.0059***	0.0003	0.0079***	0.0022**
	(0.0005)	(0.0005)	(0.0009)	(0.0008)
2nd RD/employment quintile	-0.0020***	-0.0044***	-0.0018*	0.0048***
	(0.0006)	(0.0006)	(0.0007)	(0.0009)
Middle RD/employment quintile	-0.0002	-0.0044***	0.0017*	-0.0048***
	(0.0006)	(0.0007)	(0.0008)	(0.0009)
4th RD/employment quintile	0.0217***	0.0179***	0.0097***	0.0309***
	(0.0007)	(0.0008)	(0.0009)	(0.0010)
Top RD/Employment quintiles	0.0215***	0.0203***	0.0284***	0.0340***
	(0.0008)	(0.0010)	(0.0011)	(0.0013)
Individual fixed effect	0.9637***	1.0011***	1.0064***	1.0368***
	(0.0004)	(0.0004)	(0.0005)	(0.0005)
Establishment fixed effect	0.9707***	0.9954***	0.9854***	1.0259***
	(0.0008)	(0.0009)	(0.0011)	(0.0011)
r2_a	0.964	0.963	0.957	0.956
Nround	581300	538200	398600	389000

Panel B: Sample of individuals observed at least 4 times

	1992	1997	2002	2007
Years of Education	0.0740***	0.0772***	0.0856***	0.0926***
	(0.0004)	(0.0003)	(0.0003)	(0.0004)
Experience (Age-YrsoSch -5)	0.0141***	0.0123***	0.0100***	0.0086***
	(0.0001)	(0.0001)	(0.0002)	(0.0004)
Experience squared	-0.0009***	-0.0006***	-0.0004***	-0.0004***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Dummy variable for Female	-0.3170***	-0.3150***	-0.3073***	-0.3405***
	(0.0023)	-0.0018	(0.0026)	(0.0066)
Dummy variable for Nonwhite	-0.0881***	-0.0727***	-0.0714***	-0.0662***
	(0.0023)	(0.0018)	(0.0017)	(0.0024)
Interaction female with Experience	0.0017***	0.0041***	0.0052***	0.0074***
	(0.0002)	(0.0003)	(0.0005)	(0.0009)
Interaction female with Experience^2	0.0002***	0.0001***	-0.0000	-0.0000
	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Interaction female with Nonwhite	0.0485***	0.0260***	0.0057	-0.0035
	(0.0039)	(0.0030)	(0.0029)	(0.0040)
Log number workers in firm	0.0070***	0.0060***	0.0048***	0.0053***
	(0.0006)	(0.0005)	(0.0005)	(0.0007)
Log of workers in establishment	0.0642***	0.0620***	0.0625***	0.0691***
	(0.0009)	(0.0007)	(0.0007)	(0.0010)
Mean yrs of schooling	0.0034*	-0.0001	-0.0048***	-0.0086***
	(0.0014)	(0.0011)	(0.0011)	(0.0014)
Mean yrs of experience	-0.0040***	-0.0043***	-0.0064***	-0.0051***
	(0.0002)	(0.0002)	(0.0002)	(0.0003)
Mean percent female	-0.0048	0.0060	0.0648***	0.0644***
	(0.0057)	(0.0046)	(0.0046)	(0.0065)
Mean percent nonwhite	0.0007	0.0653***	0.0831***	0.1099***
	(0.0073)	(0.0057)	(0.0055)	(0.0074)

Ln capital structure per worker	-0.0017	0.0010	0.0020***	0.0019*
	(0.0012)	(0.0006)	(0.0006)	(0.0008)
Ln equipment per worker	0.0043**	-0.0002	-0.0024**	-0.0001
	(0.0013)	(0.0008)	(0.0008)	(0.0011)
Lowest RD/employment quintile	0.0035	-0.0039*	0.0108***	0.0061*
	(0.0024)	(0.0019)	(0.0025)	(0.0027)
2nd RD/employment quintile	-0.0043	-0.0051*	0.0007	0.0132***
	(0.0030)	(0.0024)	(0.0021)	(0.0031)
Middle RD/employment quintile	-0.0112**	-0.0282***	-0.0051*	-0.0238***
	(0.0036)	(0.0028)	(0.0024)	(0.0032)
4^{th} RD/employment quintile	0.0237***	-0.0071*	0.0183***	0.0284***
	(0.0041)	(0.0031)	(0.0032)	(0.0039)
Top RD/Employment quintiles	0.0598***	0.0307***	0.0638***	0.0582***
	(0.0053)	(0.0044)	(0.0037)	(0.0051)
Individual fixed effect	0.8764***	0.9851***	1.0290***	1.1014***
	(0.0022)	(0.0017)	(0.0016)	(0.0022)
Establishment fixed effect	0.8885***	0.9832***	1.0099***	1.0906***
	(0.0034)	(0.0027)	(0.0026)	(0.0036)
r2_a	0.881	0.931	0.939	0.907
Nround	55900	55900	55900	55900

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., 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 (September), 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. NBER Working Paper no. 13043 (April), 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(3) pp. 673697.

Barth, Erling & Alex Bryson & James C. Davis & Richard Freeman, 2014. "It's Where You Work: Increases in Earnings Dispersion across Establishments and Individuals in the U.S," NBER Working Papers 20447, National Bureau of Economic Research, Inc.

Brown, Charles & James Medoff 1989 "The Employer Size-Wage Effect", *Journal of Political Economy, vol 97(5):1027-1059*

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.

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 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. NBER Working Paper no. 13351 (August), National Bureau of Economic Research, Cambridge, MA.

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

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

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, IL: The University of Chicago Press.

Lise, Jeremy, Costas Meghir & Jean-Marc Robin, 2013. "Mismatch, Sorting and Wage Dynamics," NBER Working Papers 18719, National Bureau of Economic Research, Inc.

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.

Mendes, R., G. van den Berg, M. Lindeboom, (2011) "An empirical assessment of assortative matching in the labor market", Labor Economics, forthcoming

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.

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

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, IL: University of Chicago Press.

Rastogi, Sonya and Amy O'Hara, 2012. "2010 Census Match Study." U.S. Census Bureau Report.[14]

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

Appendix: **Sorting of workers across establishments**

This section relies on the full decomposition of equation (3). Results are shown in the core of Table A (the first line and column shows variances) The second line shows the regression coefficients of the observable establishment wage premium, calculated as Z $z_{jt}d_t + Z_jD$, i.e. both time varying, such as capital intensity, and constant characteristics,, such as industry and region, from equation (3), regressed separately on the factors of each column. We find that workers with higher unobserved fixed effects tend to work in high paying establishments. Also, the sum of the individual observables, $x_{ijt}\beta_t + X_iB$, is positively associated with high paying establishments. Split onto certain factors, we find that different types of workers sort differently. One more year of education means an increase of .018 log points in establishment earnings. Women and older workers tend to work in establishments that pay less. Sorting into unobservable establishment characteristics is different. There is negative assortative matching on unobservables (note however, as has been recognized before, (Andrews et al 2008) uncorrelated measurement errors will induce a strong negative correlation between the estimated unobservable along these dimensions, see Abowd et al (2014) for earlier results, and Lise, Meghir and Robin (2013) for a critical discussion). However, individuals with positive individual observable characteristics tend to sort into high paying establishments.

Table A Regression coefficients of establishment observables and unobservables on individual characteristics

		Unobserved Individual Fixed effect	Individual measured Characteristics	Yrs of Schooling	Female	Experience	Non-white
	Variance	0.167	0.096	5.542	0.214	118.549	0.186
Establishment Observables Z,	0.033	0.021	0.108	0.018	-0.028	-0.0002	0.024
Establishment unobservable Fixed effect	0.042	-0.131	0.398	0.003	-0.027	0.0003	-0.016

Note: the table shows the regression coefficient of simple regression of the establishment observable or unobservable on the individual characteristics. Each cell is a coefficient from separate regression.

‘