BLS WORKING PAPERS

U.S. Department of Labor
U.S. Bureau of Labor Statistics
Office of Employment and Unemployment Statistics

Establishment Wage Differentials

Julia I. Lane, The National Opinion Research Center (NORC)
Laurie A. Salmon, U.S. Bureau of Labor Statistics
James R. Spletzer, U.S. Bureau of Labor Statistics

Working Paper 403
March 2007

Establishment Wage Differentials

March 2007

Abstract

Economists have long known that individual wages depend on a combination of employee and employer characteristics, as well as the interaction of the two. Although it is important to understand how employee and employer characteristics are related to wages, little is known about the magnitude and relation of these wage effects. This is primarily due to the lack of microdata which links individuals to the establishments where they work, but also due to technical difficulties associated with separating out employee and employer effects. This paper uses data from the Occupational Employment Statistics program at the Bureau of Labor Statistics that permit both of these issues to be addressed. Our results show that employer effects contribute substantially to earnings differences across individuals. We also find that establishments that pay well for one occupation also pay well for others. This paper contributes to the growing literature that analyzes firms' compensation policies, and specifically the topic of employer effects on wages.

Julia I. Lane
The National Opinion Research Center (NORC) and IZA

Laurie A. Salmon
U.S. Bureau of Labor Statistics

James R. Spletzer
U.S. Bureau of Labor Statistics

We are very grateful to many of our colleagues and to participants at various conferences and seminars for their useful comments on previous drafts of this paper. The views in this paper are those of the authors and do not necessarily reflect the views of the U.S. Bureau of Labor Statistics.

I. Introduction

Economists have long known that individual wages depend on a combination of employee and employer characteristics, as well as the interaction of the two. Although it is important to understand how employee and employer characteristics are related to wages, little is known about the magnitude and relation of these wage effects. This is primarily due to the lack of microdata which links individuals to the establishments where they work, but also due to technical difficulties associated with separating out employee and employer effects. In this paper, we use microdata from the Occupational Employment Statistics program at the Bureau of Labor Statistics that permit both of these issues to be addressed. These data contain information from more than half a million establishments, in all sectors of the economy, with wages reported for over 34 million individuals in more than 800 occupations. This paper contributes to the growing literature that analyzes the impact of firms' compensation policies, and specifically the topic of employer effects on wages.

The main contribution of this paper is the empirical estimates of how wages are influenced by the establishment at which the individual works. Our decomposition of wages into employee and employer effects is based on similar work by Groshen (1991b), which uses OLS regressions to partition the sum of squares of wages into worker and establishment components. Our results show that employer effects contribute substantially to earnings differences -- the results from our basic model show that controlling for detailed occupation, establishment dummies account for over one-fifth of wage variation. We modify Groshen's decomposition and show that these large employer effects can be only partially explained by observable characteristics such as the location, size, age, and industry of the establishment.

In order to examine the breadth of the establishment wage differential across occupations, we calculate the correlations of occupational wages within establishments. Our results are striking -- we find that establishments that pay well for one occupation also pay well for others. Even after controlling for observable establishment characteristics, we find positive wage correlations within establishments for occupations that are closely related, as well as for occupations that one would not expect to be closely related in the production process. This empirical finding has interesting implications for theories that attempt to explain the source of establishment wage differentials.

II. Background and Literature Review

IIa) Empirical Estimates of Establishment Wage Differentials

Establishment wage differentials (EWDs) are defined as the wage premium, controlling for occupation and individual characteristics, that is common to all individuals in an establishment. While economists have known about EWDs since the studies of employer wage policies in the 1940s and 1950s -- see the literature review in Segal (1986) -- it is only recently with the advent of large linked employer-employee micro-databases that systematic statistical analyses of establishment wage differentials have been conducted. The empirical strategy used by almost all of these recent studies has been to define EWDs as the percentage of individual wage variation accounted for by adding establishment indicators to a regression that already includes controls for occupation and worker characteristics.

Groshen (1991b) is the seminal article in this modern literature. Using data for six manufacturing industries from the Bureau of Labor Statistics' Industry Wage Surveys, Groshen decomposed earnings variation into occupational and establishment differentials as well as the interaction between the two. She found that establishments contribute substantially to earnings differences -- when controlling for occupation, establishment wage differentials account for a sizeable amount of individual wage variation, ranging from 12 percent in the cotton and man-made textiles industry to 58 percent in the industrial chemicals industry.

Groshen's methodology and basic findings have been replicated with other data in recent studies. Using data from 241 establishments responding to the Bureau of Labor Statistics' White Collar Pay Survey, and controlling for individual worker characteristics, Bronars and Famulari (1997) find that 18 percent of individual wage variation is due to establishment wage differentials. Extending this work to provide a comparison of the United States and Denmark, Bronars, Bingley, Famulari, and Westergard-Nielsen (1999) report that 20 percent of variation in Danish while collar pay and 36 percent of variation in Danish blue collar pay is attributable to establishment wage differentials. Using data on 50,000 managerial positions in 39 companies, and controlling for job characteristics and job requirements (as measured by Hay points), O'Shaughnessy, Levine, and Cappelli (2000) find that 8 to 9 percent of individual wage variation is due to firm wage differentials. Finally, a study of the Brazilian and Chilean labor markets, Mizala and Romaguera (1998) report that 7 to 9 percent of Brazilian wage variation and 6 to 18 percent of Chilean wage variation can be attributed to firm wage differentials.

The studies just cited use cross-sectional data with multiple individuals per establishment (or firm), and report estimates of EWDs controlling for observed heterogeneity across individuals. It is natural to wonder whether these estimated EWDs might be measuring differences in average worker skill across establishments, which would result from a sorting of individuals into establishments based on characteristics unobserved by the econometrician. Evaluating this hypothesis requires panel data with multiple observations per individual and multiple individuals per establishment. Abowd, Kramarz, and their coauthors show that firm wage differentials in France account for 25 percent of wage variation conditional on observed worker characteristics, and account for 19 percent of wage variation conditional on both observed and unobserved worker heterogeneity.[1] These results demonstrates that using longitudinal microdata to account for unobserved heterogeneity diminishes but does not remove the estimated employer effect on wages.

IIb) Theoretical Explanations for Establishment Wage Differentials

Groshen (1991a) is the classic reference regarding theoretical explanations for establishment wage differentials. She proposes and evaluates five explanations for why individual wages vary among employers. These explanations for establishment wage differentials can also be found in the somewhat older and more well established industry wage differentials literature.[2]

[1] The statistics presented in this paragraph are from personal communications with John Abowd. He has graciously provided us with the R-squareds from exact solutions, instead of the R-squareds that are based upon approximations and are reported in Abowd, Kramarz, and Margolis (1999). For the record, there is nothing wrong with the approximations. Differences between the approximations and the exact solutions lies in the fact that insufficient computing capacity for analysis of the French data did not allow for the inclusion of enough terms in the approximation to get the approximate solution close to the full least squares solution. The paper by Abowd, Finer, and Kramarz (1999) did all the calculations using the same approximations with data from the State of Washington, without computer constraints, and the R-squareds based on the approximations are fine. See Abowd and Kramarz (1999) for further details.

[2] Key references that have influenced the industry wage differentials literature are Dickens and Katz (1987), Katz and Summers (1989), Krueger and Summers (1988), and Murphy and Topel (1987).

The first explanation is that of labor quality, where employers systematically sort workers by ability as predicted by team production models. Groshen offers two key reasons for why the sorting model is not the sole source of establishment wage differentials. First, EWDs are estimated conditional on controls for occupation, and Groshen argues that detailed occupational information proxies quite well for standard human capital variables. Similarly, industry wage differentials are estimated conditional on human capital controls, and these differentials still exist after controlling for unobserved individual ability in a longitudinal analysis. Second, it is difficult to reconcile the sorting explanation with the finding that establishment and industry wage differentials apply to all occupations.

A second explanation for the existence of establishment wage differentials is that of compensating differentials. Compensating differentials are defined as a wage premium paid to workers to compensate them for undesirable working conditions. This explanation is problematic since the risk of injury is occupation specific, rather than applying to all workers in the establishment. Furthermore, the industry wage differentials literature has empirically examined and rejected the hypothesis of compensating differentials as an explanation for the wage differentials.

A third explanation for the existence of establishment wage differentials is that costly information may generate random variation in wages across employers. For example, employers may profit from individuals who find it costly to search for alternative wage offers, or employers who hire infrequently may not have adjusted their pay structure since their last hiring cycle. Groshen (1991a) rejects this explanation based on evidence that employer wage differentials are persistent.

The fourth explanation proposed by Groshen (1991a) for the existence of establishment wage differentials is efficiency wages. Efficiency wage theories, particularly those that emphasize morale, loyalty, and teamwork, can explain why workers in all occupations receive the establishment wage premium. With efficiency wages, differences across establishments resulting from a variety of factors such as monitoring costs, turnover costs, or managerial tastes generates the heterogeneity necessary to observe establishment specific pay policies. Unfortunately, there is little, if any, direct empirical evidence on the relationship between efficiency wages and establishment wage differentials.

A fifth explanation is a model where wage variation across employers results from workers bargaining over rents, or employers sharing profits with employees for other reasons. These models can generate the result that the establishment wage premium covers all occupations. However, the bargaining models are difficult to evaluate, especially their applicability outside the union sector. Groshen finds some support for rent sharing models, citing the empirical literature which tends to show a positive relationship between an individual's wage and the employer's or the industry's profits.[3]

The literature on employer-size wage differentials also evaluates similar explanations regarding why the wages of individuals are associated with the establishment where they work (see Oi and Idson, 1999, for a recent survey). Briefly, the evidence from this literature suggests that theories based on compensating differentials, union avoidance, monitoring, and rent sharing accruing from product market power contribute little to explaining the employer-size wage differential. Sorting is a more likely possibility: Brown and Medoff (1989) find that labor quality variables reduce the simple size coefficients by roughly one-half, and controlling for unobserved labor quality in a longitudinal fixed effects regression reduces the size coefficients by a further five to forty-five percent. Even so, there remains a significant size effect after controlling for both observed and unobserved labor quality. Albæk, Arai, Asplund, Barth, and Madsen (1998) also find that the sorting of workers on unobserved characteristics does not explain the estimated size effect. Troske (1999) uses linked employer-employee microdata that allows him to evaluate explanations that can not be analyzed using most databases. He finds that more skilled workers tend to work together, as predicted by team production models, and this matching reduces the employer-size wage premium by approximately 20 percent. However, Troske concludes that a large and significant employer-size wage premium still exists and remains unexplained.

A recent and comprehensive analysis of employer effects on wages is provided by Abowd and Kramarz (1999). Building on previous work in Abowd, Kramarz, and Margolis (1999) and Abowd, Finer, and Kramarz (1999), this study decomposes estimates of a simply estimated employer differential into components that are due to unobserved individual

---

[3] Hildreth and Oswald (1997) is a recent reference documenting the rent sharing hypothesis. However, Margolis and Salvanes (2000) present evidence that suggests that a sizable portion of the positive correlation between firm profits and worker earnings is due to interactions between the unobservable characteristics of the firm's workforce and the bargaining power of workers at different stages of the business cycle.

heterogeneity and unobserved firm heterogeneity. Using data for both France and the United States, Abowd and Kramarz find that 45 to 50 percent of the “raw” industry wage differential is due to unobserved firm heterogeneity, and 71 to 76 percent of the “raw” firm size wage differential is due to unobserved firm heterogeneity. While the sources of the unobserved firm heterogeneity remain unknown, these empirical estimates document that employer effects on wages exist. This type of analysis based on linked employer-employee microdata is bringing the topic of employer effects on wages back to the forefront of the theoretical and the empirical literature.

III. The Wage Decomposition Methodology

Our empirical analysis is based on the methodology used by Groshen (1991b). We have a measure of log wages W_{iej} for individual "i" in establishment "e" in occupation "j." We want to decompose the variation in wages into components attributable to occupational differentials, establishment differentials, and differences across individuals. Following Groshen, we estimate the following four regressions:

(Occ) $\quad W_{iej} = \mu + OCC_j\alpha + \varepsilon_{iej},$

(Est) $\quad W_{iej} = \mu + EST_e\beta + \varepsilon_{iej},$

(Main) $\quad W_{iej} = \mu + OCC_j\alpha + EST_e\beta + \varepsilon_{iej},$

(Cell) $\quad W_{iej} = \mu + OCC_j\alpha + EST_e\beta + (OCC_j{*}EST_e)\gamma + \varepsilon_{iej}.$

In these regressions, OCC_j is a vector of dummy variables indicating the occupation, EST_e is a vector of dummy variables indicating the establishment, and $(OCC_j{*}EST_e)$ is a vector of dummy variables indicating an occupational-establishment job cell.

This wage decomposition partitions the sum of squares of wages into its various components. As Groshen (1991b) mentions, this statistical technique avoids imposing structure on unbalanced data. The OES microdata are unbalanced, with a different number of workers across occupations, and a different number of occupations across establishments. The R-squareds from each of the four regressions are the key to the decomposition (we will not report the regression coefficients α, β, or γ). We notationally define these R-squareds as R^2_{Occ}, R^2_{Est}, R^2_{Main}, and R^2_{Cell}.

As seen from the first three regressions above, log wages are regressed on vectors of occupation and establishment indicators separately, and then on both sets of indicators together (the main effects model). The marginal contribution of establishment indicators to the main effects model, relative to the regression with just occupation indicators, measures the portion of wage variation associated unambiguously with the establishment indicators. This is calculated as (R^2_{Main} - R^2_{Occ}). Similarly, the marginal contribution of occupation indicators is calculated as (R^2_{Main} - R^2_{Est}), and measures the portion of wage variation associated unambiguously with the occupation indicators.

The explanatory power of occupation and establishment together in the main effects model does not necessarily equal the sum of the marginal contributions to the main effects model from the establishment indicators and from the occupation indicators. This difference, which is measured as ($R^2_{Est} + R^2_{Occ} - R^2_{Main}$), is referred to as the "joint" explanatory power of occupation and establishment. This joint contribution is non-zero if there is any sorting of occupations across establishments. Positive sorting occurs if high wage occupations are concentrated in establishments with high wage premiums ($R^2_{Est} + R^2_{Occ} > R^2_{Main}$), whereas negative sorting occurs if high wage occupations are concentrated in establishments with lower wage premiums ($R^2_{Est} + R^2_{Occ} < R^2_{Main}$). The existing literature -- Groshen (1991b) and Groshen and Levine (1998) -- has found positive sorting between occupational wage differentials and establishment wage differentials.

In the fourth regression above, the job cell interactions measure the wage premium paid to a particular occupation in a particular establishment above or below the wage premium predicted by the occupational and the establishment differentials. The relative contribution of the job cells in our wage decomposition is measured as (R^2_{Cell} - R^2_{Main}). The explanatory power of job cells capture what Groshen and Levine (1998) refer to as the "internal (wage) structure effect." In a wage regression, the job cells can reflect many factors. For example, the initial phases of an establishment's production process may resemble the average in the industry, but the finishing process may require workers of higher than average ability. Another example may be that the wage profile in the establishment is tilted, either because of on-the-job training given to entry-level workers, or as a result of deferring wages in order to offer workers incentives not to shirk. The job cell effects could also reflect differences in occupational tenure across establishments.

The final contribution to wages is the individual contribution. This is measured as $(1 - R^2_{Cell})$, and is the portion of the total sum of squares of wages that can not be explained by occupation and establishment indicators. This individual contribution is undoubtedly due to unobserved wage effects from gender, education, tenure, or other individual attributes that are not captured by the interactions of the occupation and establishment indicators.

In summary, we estimate four regressions of log wages on various combinations of occupation and establishment dummy variables, and we focus on the R-squareds from these four regressions. Simple comparisons of these R-squareds provide information on occupational and establishment wage differentials, the degree of occupational sorting across establishments, the importance of employer specific wage structures, and the importance of unobserved individual heterogeneity (controlling for occupation and establishment).

IV. The Data

We use microdata from the Occupational Employment Statistics (OES) program at the Bureau of Labor Statistics (BLS). The OES is an annual mail survey measuring occupational employment and wages by geographic area and by industry. Approximately 400,000 establishments are surveyed each year. The OES survey covers all full-time and part-time wage and salary workers in nonfarm industries. The survey does not cover the self-employed, owners and partners in unincorporated firms, household workers, or unpaid family workers. In 1996 the OES program began collecting wage data along with occupational employment data in every State. The survey is designed as a three-year sample, with one-third of both the certainty and non-certainty strata sampled each year.

We use the 1996 and 1997 microdata in our analysis. Our sample has 573,586 establishments with no imputations of wage or employment data.[4] These establishments

[4] The response rate for the OES survey is 78 percent (thus we have survey responses from roughly 624,000 of the 800,000 sampled establishments). The remaining sample reduction is to exclude the establishments that report employment or wage data for some but not all occupations. For example, if the human resources clerk filling out the survey does not have wage data for upper management, the clerk may report all employment data but not all wage data.

cover 34,453,430 workers categorized by occupation and wage interval. We also have information on the location, industry, size, and age of each establishment.

The OES survey asks establishments to fill out the elements of a matrix, where occupations are listed on the rows and various wage ranges are listed in the columns. For each occupation, respondents are asked to report the number of employees paid within specific wage intervals. An example of the OES survey form, with many of the occupations omitted for presentation purposes, is given in Figure 1. Separate OES survey forms are designed for each industry group, and list the occupations that are typical in the industry. Survey forms contain between 50 and 225 OES occupations, depending on the industry classification and size class of the sampled establishments. If an occupation is not listed on a survey form, the respondent is asked to include the information on a supplemental page. To reduce paperwork and respondent burden, no survey form contains every OES occupation.

The occupational data in the 1996 and 1997 OES survey are based on the 1980 Standard Occupational Classification (SOC) System. Occupations are classified based upon work performed, skills, education, training, and credentials. There are 824 detailed occupations in our OES microdata. In some of our analysis, we aggregate these 824 detailed (five-digit) occupational codes into seven major (one-digit) occupations: Management, Professional, Sales, Clerical, Services, Agricultural, and Production.

As seen in Figure 1, the wage information provided by establishments in the OES survey is recorded in intervals for either hourly or annual rates of pay. The actual values we use for these intervals are the mean wage of all workers within the interval as computed from the National Compensation Survey for that year.[5] In the following section and in the appendix of this paper, we discuss the econometrics and the empirical consequences of wage data reported as interval means. All of the wages used in our analysis are measured, in real terms, as the natural logarithm of hourly rates of pay.[6]

---

[5] The interval mean for the bottom interval may vary for states with a minimum wage above the federal minimum. The interval mean for the top interval is set in nominal terms at \$60.01. This upper wage interval contains 0.7% of the individuals in our sample (244,727 / 34,453,430). We have found that the results from our wage decomposition are not sensitive to the point estimate used for this upper interval: the establishment effects we report in Table 1 are 20.86% using the point estimate of \$60.01, and would be 20.78% using a point estimate of \$70.01 and 20.69% using a point estimate of \$80.01.

[6] Given that the wage data can be reported as either annual or hourly, there is a concern that the establishment wage differentials could reflect hours differences across establishments. The example of banking comes to mind: a bank with "bankers' hours" may have tellers working six hours per day, whereas a full service bank may have tellers working eight hours per day. Our estimated EWDs could be affected if earnings for occupations

The obvious strengths of the OES microdata for economic analysis are the sample size and the level of occupational detail. Specifically, there are more than half a million establishments in our sample, covering over 34 million individuals in more than 800 occupations. As such, the OES data can be viewed as a proxy for matched employer-employee microdata. The second strength of the OES is the employer-reported occupational data. Although the dataset contains no information regarding the worker's demographic characteristics (such as age, race, or gender) or other labor market characteristics (such as tenure, experience, or training), we note that the detailed occupational information should proxy for a worker's skills. We will return to this latter point in the discussion of our empirical estimates.

V. Empirical Wage Decompositions

Va) Basic Results

We present the results of our wage decomposition in Table 1. In the first column, we report estimates using the seven one-digit occupation measures. In the second column, we report estimates using the 824 five-digit occupation measures. The first four rows report the R-squareds from the regressions described earlier.[7] The bottom five rows report the various contributions of occupation and establishment to wage variation.

The R-squareds in the fourth row of Table 1 demonstrate that knowing an individual's occupation and workplace go a very long way to explaining individual wage variation. More than 72 percent of wage variation is explained by knowing the individual's one-digit occupation and establishment, and almost 88 percent of wage variation is explained by knowing the individual's five-digit occupation and establishment. This implies that approximately 12 percent of wage variation is left to unobserved individual heterogeneity (although we acknowledge that this is probably an underestimate because of our use of interval data).

---

with hours variation across establishments are reported on an annual basis. This potential bias should be mitigated, however, by the fact that the OES survey respondents are instructed to classify part time workers according to an hourly rate.

[7] These regressions are estimated from our microdata covering 34 million individuals. The R-squareds from a regression using 34 million individuals are identical to the R-squareds from a regression using 7,778,248 "cells" weighted by employment, where a "cell" is a wage interval within an establishment-occupation job cell.

The importance of the information contained in the detailed occupational categories becomes clear from an analysis of the first row in Table 1. In the first column, the seven one-digit occupation indicators explain more than 28 percent of wage variation. In the second column, the 824 five-digit occupation indicators explain more than 54 percent of wage variation. This empirically confirms our belief that the OES occupational data provide meaningful information about the work performed in the job, as well as the skills, education, training, and credentials of the persons performing the work.

The R-squareds in the second row illustrate that establishment indicators alone explain about half of the wage variation. This regression is of interest other than its intermediary role in our wage decomposition. Kremer and Maskin (1996) have developed an index which captures the degree to which workers with similar wages are grouped across establishments. The Kremer and Maskin segregation index is nothing more than the R-squared from a regression of individual wages on a vector of establishment dummies. Our estimate of .4955 is roughly comparable to other estimates from the United States.[8]

In the bottom half of Table 1, we report the decomposition of wage variation into its component parts. Looking at the second column, we find that 26 percent of wage variation is associated unambiguously with occupation, and 21 percent of wage variation is associated unambiguously with information on the worker's establishment. An important part of the story is the sorting between occupations and establishments -- we find that this joint contribution accounts for 29 percent of wage variation. And the final portion of the explained wage variation is the job cell contribution, which accounts for just over 12 percent of wage variation. The residual 12 percent of wage variation in the OES data is due to unobserved variation across individuals within a job cell.

[8] Davis and Haltiwanger (1991) report that 51 to 58 percent of the total variance in wages is accounted for by the dispersion in mean wages across plants. One can manipulate Groshen's (1991b) estimates in her Table 2 and conclude that the R-squareds from regressions of log wages on establishment dummies range from .17 to .75, with a simple mean across the six industries of .49. Bronars and Famulari (1997) report an R-squared of .45. The results in Lane, Lerman, and Stevens (1998) suggest that the proportion of wage variance explained by between firm variation is roughly .45. Outside the United States, Kramarz, Lolliver, and Pelé (1996) report a wage-based segregation measure for France of .38 in 1986 and .48 in 1992, and Bronars, Bingley, Famulari, and Westergard-Nielsen (1999) report an R-squared of .35 for white collar workers and .46 for blue collar workers in Denmark.

It is interesting to compare the results of our wage decomposition with the results reported by Groshen (1991b).[9] If we compute the simple average across the six industries reported by Groshen, her results fall in between the results we report in columns 1 and 2 of Table 1. For example, Groshen's estimates imply that occupation indicators account for a mean of 20 percent of wage variation, and establishment indicators account for a mean of 32 percent of wage variation. Our estimates of the occupation effect range from 15 to 26 percent, and our estimates of the establishment effect range from 21 to 36 percent. Our estimates of the joint sorting effect (14 to 29 percent), the job cell effect (8 to 12 percent), and the individual effect (12 to 27 percent) are also comparable to the means of the estimates reported by Groshen (17 percent, 10 percent, and 22 percent, respectively).

The estimates in Table 1 provide interesting insight into the labor market and the wage setting practices of businesses. The occupation and establishment information in the OES data explain most of the wage variation across individuals. We find, not surprisingly, that detailed information on the individual's occupation explains a sizable amount of wage variation. Building on a small but growing literature, we also find substantial establishment wage differentials.

Vb) Sensitivity Analysis

The R-squared of .8798 in Table 1 is unusually high if one were to compare it to most earnings regressions based on worker surveys. We are not the first to find such a high R-squared when employers are included: Groshen (1991b, page 869) finds that "occupation and establishment identity alone can explain over 90 percent of wage variation among blue-collar workers." It is interesting to note that this high R-squared is achieved despite that fact that education and other individual determinants of wages are not available, confirming that occupation is a strong proxy for these factors. This is supported by our finding that the residual individual component falls from .27 to .12 when we move from one-digit to five-digit occupation controls.

[9] We recognize that it may be conceptually difficult to compare our results (which are computed from a national sample) with Groshen's results (which are computed from six industries). One purpose of this simple comparison is to demonstrate that the results from our estimation, and in particular the high R-squareds, are similar to results from other data which use the same methodology.

However, it is possible that despite the fact that the OES survey contains some of the most detailed and accurate occupational data available in any dataset, the R-squared may be inflated for technical reasons -- the wage intervals in which the data are reported may be "too wide" relative to the wage variation within establishments. Clearly, as the occupational classifications become more detailed, or as the wage intervals become wider, the average number of wage intervals reported per job cell will decrease and the R-squareds will increase. In this section, we examine the possibility that this may be a source of bias by undertaking an extensive sensitivity analysis.

In our full sample, 62 percent of job cells contain all employment within one OES wage interval, whereas 16 percent of job cells contain employment in 3 or more OES wage intervals.[10] The average number of OES wage intervals for a job cell is 1.66. Obviously these statistics will vary by establishment size. For the largest establishments (those with more than 500 employees), 37 percent of job cells contain employment in only one OES wage interval, whereas 38 percent of job cells contain employment in 3 or more OES wage intervals. We find it interesting that for these largest establishments, the R-squareds in the job cell regressions are still quite high at .8288 (see Appendix Table 1).

The technical appendix to this paper describes an econometric framework for analyzing how collecting wage data in intervals affects the R-squareds from our wage decomposition. The essence of this framework is to simulate a distribution of individual wages which can then be collapsed to intervals corresponding to the OES wage intervals. The following is a summary of our results. If a continuous distribution of individual wages were used as the dependent variable in the regression of wages on occupation dummies, we calculate that the R-squared would be .5292 instead of the .5466 we report in Table 1. Similarly, the R-squared for the regression on establishment dummies should be .4798 instead of the .4955 reported in Table 1. And our simulation suggests that after accounting for the effect of interval means, the R-squared for the main effect regression would fall from .7552 to .7312, and the R-squared for the job cell regression would fall from .8798 to .8518. Transforming these simulated R-squareds into the occupational and establishment contributions to wage variation, our estimates of {.2597, .2869, .2086, .1246, .1202} reported

[10] See Appendix Table 1 for these statistics. The unit of analysis in Appendix Table 1 is the job cell, which is a detailed 5-digit occupation within an establishment.

in Table 1 would change to {.2514, .2778, .2020, .1206, .1482}. Each of the first four terms (the occupational effect, the joint effect, the establishment effect, and the job cell effect) falls slightly, and the residual individual effect rises from .1202 to .1482. We conclude that collecting individual wage data as intervals in an establishment survey does not distort the conclusions we draw from our wage decomposition.[11] Indeed, this evidence supports the notion that an important source of earnings variation comes from between, rather than within, establishment variation.

Vc) A Closer Examination of Establishment Wage Differentials

In column 2 of Table 1, we found that 20.9 percent of wage variation is attributable to differences across establishments. This is strong evidence for establishment wage differentials. However, these estimated EWDs might simply reflect cost of living differences across establishments in different geographical areas, or might be acting as a proxy for other characteristics such as size or industry. We explore the importance of these effects by modifying Groshen's (1991b) decomposition to include establishment level explanatory variables such as age, size, industry, and county in the right hand side of the wage regression.

Our wage decomposition is now based on five regressions, where the additional regression is:

$$\text{(Occc,X)} \qquad W_{iej} = \mu + OCC_j\alpha + X_e\delta + \varepsilon_{iej},$$

The components of X_e are dummy variables for industry, county, age, and size.[12] The R-squared from this fifth regression is notationally defined as $R^2_{Occ,X}$. Since these explanatory variables are linear combinations of the establishment dummies, we can decompose the establishment contribution of the wage decomposition into two pieces: the explained and the

[11] We present some additional evidence in support of this conclusion. We took a cross-section of Unemployment Insurance wage records (described in Burgess, Lane, and Stevens, 2000) and regressed a continuous measure of annual earnings on firm dummies, and then ran a similar regression where the dependent variable has been recoded into point estimates corresponding to the OES hourly wage intervals. The R-squareds are quite similar: .55 with the continuous measure, and .57 with the interval measure.

[12] We considered allowing age and size to be continuous variables rather than a vector of dummy variables, but we adopted our dummy variable approach for several reasons. Intuitively, we maintain the spirit of "occupation first" followed by "establishment conditional on occupation" that is implicit in Groshen's (1991b) wage decomposition. Furthermore, the joint contribution is unaffected, which says that the sorting effect between occupations and establishments has not changed. And finally, we note that Abowd, Kramarz, and Margolis (1999) recoded their continuous size variable into dummy variables.

unexplained contribution. We define the explained component of the establishment effect as ($R^2_{Occ,X}$ - R^2_{Occ}), and the unexplained component of the establishment effect as (R^2_{Main} - $R^2_{Occ,X}$). These two components sum to the total establishment effect in table 1, which is calculated as (R^2_{Main} - R^2_{Occ}).

The wage decompositions controlling for the effects of observable establishment characteristics are presented in Table 2. In column 1, we present the wage decomposition controlling for any county effects including cost of living differences that are common within counties. These county controls account for one-fifth of the estimated establishment wage differentials (.0418/.2086), and thus local area differences explain some of why wages vary across establishments. Similarly, in columns 2 through 5 of Table 2, we conclude that age, size, and industry can each only explain a small portion of why wages vary across establishments.[13] When we control for all observable effects together in column 6 of Table 2, we account for half of the estimated establishment wage differentials. We conclude that establishment wage differentials can be only partially explained by observable establishment characteristics, and thus establishment wage differentials are an important explanation for why wages vary across individuals.

Vd) Further Empirical Results

Many of the explanations put forward for the existence of employer effects on wages vary in importance for different industries. For example, capital-labor complementarity should be more important in the goods producing industries than in the services providing industries, unionization rates vary dramatically across industries, and skill sorting should be more important in industries that produce heterogeneous output. The results presented in Table 3a show noticeable differences across major industries. Establishment wage differentials are most important in construction, mining, manufacturing, and TCPU, and are least important in public administration, FIRE, agriculture, and services. EWDs explain 37 percent of wage variation in construction, yet only 16 percent of wage variation in the

[13] Since the largest size category of 500+ accounts for about 40 percent of the workers (but only 1.4 percent of the establishments), it is a worthwhile exercise to experiment with alternative size categories. We find that the amount of the establishment effect that can be explained using observed establishment characteristics is not sensitive to the categorization of the size variable. When we break out the 500+ category into deciles and use 18 size class dummies (instead of the 9 used in Table 2), the size variable accounts for 11.7 percent of our estimated establishment effect (compared to 10.5 percent calculated from the statistics reported in Table 2).

services industry. A number of reasons for these industry differences are possible: the traditional goods producing industries are more unionized than the other sectors (with the exception of public administration), and these industries may well have greater variation in capital usage.

Interestingly, the construction and services industries are also quite different with regard to the contribution of occupational sorting: this component of the wage decomposition contributes little to variation in earnings in construction, but is quite important in services. This suggests that establishments in the construction industry bundle their workers in very similar ways, while establishments in the services industry bundle their workers very differently.

It is equally rewarding to analyze differences by establishment size. As seen in Table 3b, the importance of establishment wage differentials drops markedly and monotonically with the size of the establishment. EWDs explain 30 percent of wage variation for establishments with 2-9 employees, yet explain 16.5 percent of wage variation for the largest establishments. We also see that the percentage of the establishment effect that can be explained by observed characteristics rises with the size of the establishment. Our finding that small establishments have more variation, both total and unexplained, in their contribution to wages is consistent with the notion that small establishments are more idiosyncratic than large establishments with regard to their personnel and paysetting practices.[14]

VI. Occupational Wages Within Establishments

The empirical evidence from our wage decompositions highlights the importance of the establishment for understanding the variation of individual wages. Even after controlling for observable characteristics that vary across establishments, we find substantial evidence of establishment wage differentials. By definition, these establishment wage differentials measure the wage premium paid to all workers in the establishment, regardless of occupation. We now turn to examining the correlations of occupational wages within establishments. Our

---

[14] This conclusion mirrors the findings of Haltiwanger, Lane, and Spletzer (2006), who show that new businesses exhibit greater earnings heterogeneity than do mature businesses.

analysis here is motivated by the team production model, well described by Kremer (1993). Simply put, in this model, workers of similar skill will match together in firms -- highly skilled supervisors will work with highly skilled production workers. This reflects the complex nature of a multi-stage production process, which requires the coordinated and successful completion of distinct tasks. In many production processes, it is not possible for several low skilled workers to substitute for one high skilled worker. Empirically, this should result in a positive correlation of occupational wages within establishments.

Our analysis in this section is similar to previous work by Dickens and Katz (1987), Bronars and Famulari (1997), and Bronars, Bingley, Famulari, and Westergard-Nielsen (1999). The goal of our correlation analysis is to examine the breadth of the establishment wage differentials across occupations. For example, in a manufacturing plant, we would expect the wages of machinists and production supervisors to be positively correlated since they work side by side on the assembly line. It is less likely, however, that wages of the accountants or the janitors in this manufacturing plant would be positively correlated with the wages of the machinists and the production supervisors.

An examination of the data reveals that while the correlations across closely related occupations are quite high, supporting a team production hypothesis, correlations are also surprisingly high across unrelated occupations. In Figure 2, continuing with the example from the previous paragraph, we graph the average wages of one occupation against the average wages of another occupation in the same establishment.[15] We find, not surprisingly, that the wages of machinists and the wages of production supervisors are closely correlated (the correlation is .61).[16] We also find that the wages of accountants are positively correlated with the wages of machinists and production supervisors (the correlations are .43 and .41), and the wages of janitors are positively correlated with the wages of machinists and production supervisors (the correlations are .61 and .55). Perhaps most surprisingly, the wages of accountants are highly correlated with the wages of janitors in the same establishment (the correlation is .41).

[15] There are 47,633 manufacturing establishments with at least one worker in any of the four occupations. We have selected the 338 manufacturing establishments with at least 2 workers in each of the four occupations.

[16] The correlation coefficient is the square root of the R-squared from an OLS regression of one occupational mean wage against another occupational mean wage. For example, the R-squared from a regression of the mean wages of machinists against the mean wages of production supervisors is .37 (=.61*.61).

Consistent with our earlier analysis of establishment wage differentials, the enormous wage heterogeneity across the manufacturing establishments that is evident in Figure 2 deserves mention. For example, the establishment mean ln(wage) of accountants in this sample ranges from 2.1 to 3.9 (with a mean of 2.94 and a standard deviation of 0.26). This heterogeneity is consistent with the findings of Haltiwanger, Lane, and Spletzer (2006), who outline a model where some unobserved business "type" generates heterogeneity in establishment productivity and wages. Furthermore, our findings in Figure 2 of skill complementarity across occupations within the establishment fits quite nicely with Haltiwanger, Lane, and Spletzer's model of complementarity between the "type" of business and the skill composition of its workforce.

We investigate the relationship of occupational mean wages within establishments more formally in Table 4. For the seven major occupations, we present the correlation matrix of occupational mean wages within establishments. We present two correlations for each occupational pair. The top correlation is unadjusted for observable establishment characteristics, whereas the bottom correlation is based on individual wage data with county, age, size, and major industry means removed.

Looking at the data unadjusted for establishment characteristics, the average of the 21 off-diagonal correlations is .4614. This is very similar to the estimate of Bronars and Famulari (1997), who report a correlation of mean occupational wages between professional and nonprofessionals of .499. All these correlations in Table 4 are positive and statistically greater than zero at conventional levels of significance. This says that establishments that pay well for one occupation also pay well for all other occupations. One particularly interesting pattern is that all correlations below .4 are in the upper right corner of the table -- it would seem that the least skill matching within establishments occurs between traditional white collar occupations (managers, professionals, and sales) and blue collar occupations (services, agricultural, and production). The correlations in Table 4 are consistent with theories which predict that workers are sorted into establishments based on skill.

Just as with the wage decomposition analysis, it is possible that these correlations are biased upward by not controlling for observable characteristics of the establishment. After taking out the effects of county, age, size, and industry, it is clear that the correlations fall. The average off-diagonal correlation has fallen dramatically from .4614 to .2700. However, the correlations remain quite large, and all the correlations remain statistically greater than zero. This leads us to conclude that the unadjusted occupational mean correlations within establishments do measure cost of living differences, industry effects, or size effects to a large extent, but are also measuring establishment specific pay practices that are unobservable to the econometrician.

VII. Conclusions and Discussion

Using the wage decomposition proposed by Groshen (1991b), we have documented the magnitude of occupation and establishment wage differentials, the sorting of high wage occupations into high wage establishments, and the extent of employer specific wage structures -- the wage premium paid to particular occupations in particular establishments above or below the wage premium predicted by the occupational and the establishment differentials. Our key finding in this paper is the large effect that the establishment has on the wages of the individuals who work there. We find that controlling for detailed occupation, 21 percent of wage variation can be explained by merely knowing the individual's establishment. Accounting for observable characteristics of the employer reduces these establishment wage differentials by half. Taking our empirical analysis one step further, we have shown that the establishment's wage premium is correlated across major occupation groups within the establishment. These empirical estimates complement and enhance previous work on the topic of employer effects on wages.

One of the dominant themes running through the literature of employer effects on wages is that establishments systematically sort workers by skill. The existing empirical work finds that this sorting explains much but not all of the observed employer effects on wages. Our findings are certainly consistent with this conclusion. In our wage decomposition, merely knowing the worker's establishment explains 50 percent of the observed wage variation across individuals. Controlling for the seven one-digit occupation indicators lowers this wage variation explained by establishments to 36 percent, and controlling for five-digit occupation indicators lowers this further to 21 percent. Since our detailed occupational information proxies for the worker’s skills, we also find that controlling for skill explains much, but certainly not all, of the estimated establishment wage differentials in the raw data.

Another of the themes running through the literature is that establishment wage differentials merely proxy, at least in part, for unobserved characteristics of the establishment that are correlated with wages. Our results are consistent with this hypothesis. To the extent that differences across establishments in working conditions, cost of living, rent sharing, and capital-labor ratios can be proxied for by observable establishment characteristics such as county, age, size, and industry, we find that controlling for these characteristics lowers the estimated establishment wage differentials from 21 percent of wage variation to 10 percent.

We are now left with the question of how to explain our estimated establishment wage differentials. Any explanation we propose must simultaneously account for our finding that the establishment wage differentials are common to workers in all occupations in the establishment.

One possible explanation is that the observed differentials simply reflect differences in unobserved labor quality across establishments, and that more detailed information on individual ability and human capital would serve to eliminate the EWDs. To the extent that this explanation is true, EWDs support the sorting theory, to the extent that it is not, EWDs support variations in establishment pay practices. Testing this hypothesis is beyond the capabilities of our dataset, since we don't have information on worker characteristics such as education, age, tenure, or training. However, there are several reasons to doubt that this hypothesis is the sole explanation of the estimated EWDs. First, the work of Groshen (1991b) and Levine (1992) suggests (but does not prove) that occupation adequately controls for standard measures of human capital. In addition, work by O'Shaughnessy, Levine, and Cappelli (2000) finds that measures of skill and job characteristics do not explain much of the difference in wages across employers (although these measures of skill explain quite a bit of wage variation across individuals). The findings of John Abowd and his colleagues, who have access to longitudinal linked employer-employee microdata and are thus able to proxy for unobserved skill using person-specific dummy variables, suggest that unmeasured heterogeneity across individuals explains some but not all of the estimated employer effects on wages. And finally, it is difficult to theorize how unobserved ability and human capital could be important contributors to wage differentials across all occupations in the establishment -- such as janitors and accountants.

Another possibility is that the observed differentials reflect differences in technology or capital across establishments. Pekkarinen (2002) presents an interesting model and analysis along these lines. Recent work using establishment microdata has illustrated the striking amount of heterogeneity across establishments within narrowly defined aggregates -- see, for example, Haltiwanger, Lane, and Spletzer (2006). While we have used establishment characteristics such as age, size, and industry to proxy for such differences, it would be useful to incorporate establishment level information on inputs to (and outputs from) the production process into our analysis. However interesting and worthwhile this line of research would be, we believe it is unlikely that capital intensity or technology *per se* would produce establishment wage differentials that are common to all occupations -- again, the example of janitors and accountants comes to mind.

We believe that any explanation for the existence of establishment wage differentials will rest on a combination of theories. Empirical work from recent analysis of matched employer-employee data shows that higher skilled workers not only work together in the same establishment, but also tend to work with higher quality capital and technology -- see Doms, Dunne, and Troske (1997) and Haltiwanger, Lane, and Spletzer (2006). Modeling these basic human capital results, augmented with a theory of why human resource pay policies might differ across establishments, should show how the gains from skill sorting and capital-labor complementarities can be spread to workers in all occupations in the establishment. These thoughts are not original to us, but run through the existing literature examining why the wages of individuals are affected by their employer. There is much more to be learned from additional theoretical and empirical research.

References

Abowd, John M. and Francis Kramarz (1999). "Inter-Industry and Firm-size Wage Differentials in France and the United States." Unpublished paper, Cornell University.

Abowd, John M., Francis Kramarz, and David Margolis (1999). "High Wage Workers and High Wage Firms." *Econometrica*, pp. 251-334.

Abowd, John M., Hampton Finer, and Francis Kramarz (1999). "Individual and Firm Heterogeneity in Compensation: An Analysis of Matched Longitudinal Employer-Employee Data for the State of Washington." In *The Creation and Analysis of Employer-Employee Matched Data*, edited by John C. Haltiwanger, Julia I. Lane, James R. Spletzer, Jules J.M. Theeuwes, and Kenneth R. Troske, North-Holland Press, pp. 3-24.

Albæk, Karsten, Mahmood Arai, Rita Asplund, Erling Barth, and Erik Støjer Madsen (1998). "Measuring Wage Effects of Plant Size." *Labour Economics*, pp. 425-448.

Bronars, Stephen G. and Melissa Famulari (1997). "Wage, Tenure, and Wage Growth Variation Within and Across Establishments." *Journal of Labor Economics*, pp. 285-317.

Bronars, Stephen G., Paul Bingley, Melissa Famulari, and Niels Westargard-Nielsen (1999). "Employer Wage Differentials in the United States and Denmark." In *The Creation and Analysis of Employer-Employee Matched Data*, edited by John C. Haltiwanger, Julia I. Lane, James R. Spletzer, Jules J.M. Theeuwes, and Kenneth R. Troske, North-Holland Press, pp. 205-229.

Brown, Charles and James Medoff (1989). "The Employer Size-Wage Effect." *Journal of Political Economy*, pp. 1027-1059.

Burgess, Simon, Julia Lane, and David Stevens (2000). "Job Flows, Workers Flows, and Churning." *Journal of Labor Economics*, pp. 473-502.

Davis, Steve J. and John Haltiwanger (1991). "Wage Dispersion Between and Within U.S. Manufacturing Plants." *Brookings Papers on Economic Activity*, pp. 115-200.

Dickens, William T. and Lawrence F. Katz (1987). "Inter-Industry Wage Differences and Theories of Wage Determination." NBER Working Paper #2271.

Doms, Mark, Timothy Dunne, and Kenneth R. Troske (1997). "Workers, Wages, and Technology." *The Quarterly Journal of Economics*, pp. 253-290.

Groshen, Erica L. (1991a). "Five Reasons Why Wages Vary Among Employers." *Industrial Relations*, pp. 350-381.

Groshen, Erica L. (1991b). "Sources of Intra-Industry Wage Dispersion: How Much do Employers Matter?" *The Quarterly Journal of Economics*, pp. 869-884.

Groshen, Erica L. and David I. Levine (1998). "The Rise and Decline (?) of U.S. Internal Labor Markets." Working Paper #9819, Federal Reserve Bank of New York.

Haltiwanger, John C., Julia I. Lane, and James R. Spletzer (2006). "Wages, Productivity, and the Dynamic Interaction of Businesses and Workers." Forthcoming, *Labour Economics*.

Hildreth, Andrew K.G. and Andrew J. Oswald (1997). "Rent-Sharing and Wages: Evidence from Company and Establishment Panels." *Journal of Labor Economics*, pp. 318-337.

Katz, Lawrence F. and Lawrence H. Summers (1989). "Industry Rents: Evidence and Implications." *Brookings Papers on Economic Activity*, pp. 209-275.

Kramarz, Francis, Stéfan Lollivier, and Louis-Paul Pelé (1996). "Wage Inequalities and Firm-Specific Compensation Policies in France." *Annales D'Economie Et De Statistique*, pp. 369-386.

Kremer, Michael (1993). "The O-Ring Theory of Economic Development." *The Quarterly Journal of Economics*, pp. 551-575.

Kremer, Michael and Eric Maskin (1996). "Wage Inequality and Segregation by Skill." NBER Working Paper #5718.

Krueger, Alan B. and Lawrence H. Summers (1988). "Efficiency Wages and the Inter-Industry Wage Structure." *Econometrica*, pp. 259-294.

Lane, Julia, Robert I. Lerman, and David Stevens (1998). "Employers, Jobs, and the Dynamics of Earnings Inequality." Unpublished paper, American University.

Levine, David I. (1992). "Can Wage Increases Pay for Themselves? Tests with a Production Function." *Economic Journal*, pp. 1102-1115.

Margolis, David N. and Kjell G. Salvanes (2000). "Do Firms Really Share Rents With Their Workers?" Unpublished paper presented at the Society of Labor Economists Conference.

Mizala, Alejandra and Pilar Romaguera (1998). "Wage Differentials and Occupational Wage Premia: Firm-Level Evidence for Brazil and Chile." *Review of Income and Wealth*, pp. 239-257.

Murphy, Kevin M. and Robert H. Topel (1987). "Unemployment, Risk, and Earnings: Testing for Equalizing Wage Differences in the Labor Market." In *Unemployment and the Structure of Labor Markets*, edited by Kevin Lang and Jonathan S. Leonard, Basil Blackwell, pp. 103-140.

Oi, Walter Y. and Todd L. Idson (1999). "Firm Size and Wages." In *Handbook of Labor Economics*, edited by Orley Ashenfelter and David Card, North-Holland Press, pp. 2165-2214.

O'Shaughnessy, K.C., David I. Levine, and Peter Cappelli (2000). "Changes in Managerial Pay Structures 1986-1992 and Rising Returns to Skill." NBER Working Paper #7730.

Pekkarinen, Tuomas (2002). "Complexity, wages, and the O-ring production function: evidence from Finnish panel data." *Labour Economics*, pp. 531-536.

Segal, Martin (1986). "Post-Institutionalism in Labor Economics: The Forties and Fifties Revisited." *Industrial and Labor Relations Review*, pp. 388-403.

Troske, Kenneth R. (1999). "Evidence on the Employer Size-Wage Premium from Worker-Establishment Matched Data." *The Review of Economics and Statistics*, pp. 15-26.

Technical Appendix: OES Wages and Interval Econometrics

The OES survey collects employee wage data in intervals. As is evident from Figure 1, this eliminates any wage heterogeneity across individuals within an interval. However, because there are multiple wage intervals for a given occupation on the OES survey form, there is still wage heterogeneity across individuals within job cells (where job cells are defined as an occupation within an establishment). In our decomposition, where we partition the total sum of squares into its components, this interval method of collecting individual wage data should reduce the residual variance attributable to individuals and thus increase the explained variance due to establishments and occupations. How severe might this problem be? In this appendix, we present an econometric framework for simulating how this data collection methodology affects the estimates from our wage decomposition.

Assume that an individual's true ln(wage) is Y_{iej}, but the data analyst observes W_{iej} -- the natural logarithm of the OES interval mean. The relationship between the observed wage and the true wage is $W_{iej}=Y_{iej}+\omega_{iej}$, where ω_{iej} measures how the individual's wage differs from the interval mean. For example, in Figure 1, wage interval "H" includes all employees earning between \$19.25 and \$24.24 per hour, and the OES interval mean for survey year 1997 is $W_{iej}=21.43$. With appropriate transformations to logarithms, ω_{iej} in this example is bounded between -.1073 and .1232. We shall refer to ω_{iej} as the "interval error."

For any vector of explanatory variables X, the R-squared that we estimate from the regression $W_{iej}=X_{iej}\beta^W+\varepsilon_{iej}$ is:

$$R_W^2 = 1 - \left[\frac{(W - X\hat{\beta}^W)'(W - X\hat{\beta}^W)}{(W - \overline{W})'(W - \overline{W})} \right].$$

But the "true" unobserved individual wage (Y_{iej}) should have been used as the dependent variable in the regression, rather than the observed interval mean (W_{iej}). The R-squared that would have been estimated from the regression $Y_{iej}=X_{iej}\beta^Y+\varepsilon_{iej}$ is:

$$R_Y^2 = 1 - \left[\frac{(Y - X\hat{\beta}^Y)'(Y - X\hat{\beta}^Y)}{(Y - \overline{Y})'(Y - \overline{Y})} \right]$$

$$= 1 - \left[\frac{(W - \omega - X\hat{\beta}^Y)'(W - \omega - X\hat{\beta}^Y)}{(\{W - \omega\} - \{\overline{W} - \overline{\omega}\})'(\{W - \omega\} - \{\overline{W} - \overline{\omega}\})} \right].$$

If we assume that $X'\omega=0$, so that $\hat{\beta}^Y = \hat{\beta}^W$,[17] then

[17] If X is a matrix of establishment indicators, a sufficient condition for $X'\omega=0$ is that the mean of the interval error is zero for every establishment. While one can come up with examples of certain establishments where this sufficient condition may not be true, we have no reason to believe that $X'\omega\neq0$ in the overall sample.

$$R_Y^2 = 1 - \left[\frac{(W - X\hat{\beta}^W)'(W - X\hat{\beta}^W) - 2\omega' W + \omega'\omega}{(W - \overline{W})'(W - \overline{W}) - 2(\omega - \overline{\omega})'(W - \overline{W}) + (\omega - \overline{\omega})'(\omega - \overline{\omega})} \right].$$

Because the point estimates for the interval wages are computed as the mean of the underlying wage distribution for each interval using data from the Employment Cost Index, we know that $\overline{\omega}$=0 for each interval.[18] Since W is a fixed value within each interval, it is straightforward to show that W'ω=0, and thus

$$R_Y^2 = 1 - \left[\frac{(W - X\hat{\beta}^W)'(W - X\hat{\beta}^W) + \omega'\omega}{(W - \overline{W})'(W - \overline{W}) + \omega'\omega} \right].$$

Given our assumptions, one can show that $R_Y^2 < R_W^2$. Therefore, when using interval means rather than the true unobserved wages, the R-squareds that we obtain from our regressions overstate the contribution of occupation and establishment indicators to wage variation, and thus understate the residual contribution of unobserved individual heterogeneity.

We now describe our simulation exercise that is based on this econometric framework. We have simulated a ln(wage) for 34,453,430 individuals from a normal distribution with mean 2.5133 and standard deviation 0.5446 (this mean and standard deviation are reported in the footnotes to Table 1). We then compute the corresponding wage level, define the interval wage corresponding to the OES intervals reported in Figure 1, and also define the interval error ω_{iej} as the difference between the individual's true wage and the natural logarithm of the interval wage. After forcing the interval error ω_{iej} to be mean zero within intervals, we calculate all the terms necessary to compare R_Y^2 and R_W^2 as defined in the formulas above.

We report in Table 1 that the R-squared for the regression of wages on occupation dummies is .5466. If the individual wages instead of interval means were used as the dependent variable in the regression, we calculate that this R-squared should be .5292.[19] Similarly, the R-squared for the regression on establishment dummies should be .4798 instead of the .4955 reported in Table 1. And our simulation suggests that after accounting for the effect of interval means, the R-squared for the main effect regression would fall from .7552 to .7312, and the R-squared for the job cell regression would fall from .8798 to .8518.

Transforming these simulated R-squareds into the occupational and establishment contributions to wage variation, our estimates of {.2597, .2869, .2086, .1246, .1202} reported in Table 1 would change to {.2514, .2778, .2020, .1206, .1482}. Each of the first four terms (the occupational effect, the joint effect, the establishment effect, and the job cell effect) falls

[18] This statement is not true for the uppermost interval, where the OES interval mean is set in nominal terms at $60.01. In our econometric framework, we censor the simulated wage distribution at this value and thus by definition impose a zero mean for this interval.

[19] The original R-squared of .5466 is calculated as [1-(4,633,689/10,219,022)]. The simulated R-squared of .5234 is calculated as [1-(4,633,689+335,893)/(10,219,022+335,893)], where ω'ω=335,893.

slightly, and the residual individual effect rises from .1202 to .1482. We conclude that having individual wage data calculated as interval means does not distort the conclusions we draw from our wage decomposition.

Figure 1: Example of OES Survey Form
Nonmetallic Minerals and Metal Mining Industries

Occupation Code and Title			Number of Employees in Selected Wage Ranges											Number of Employees
			A	B	C	D	E	F	G	H	I	J	K	
		Hourly	Under $6.75	$6.75-8.49	$8.50-9.99	$10.00-11.24	$11.25-13.24	$13.25-15.74	$15.75-19.24	$19.25-24.24	$24.25-43.24	$43.25-$60.00	$60.01 and over	
		Annual	Under $14,040	$14,040-17,659	$17,660-20,779	$20,780-23,999	$23,400-27,559	$27,560-32,759	$32,760-40,039	$40,040-50,439	$50,440-89,959	$89,960-124,820	$124,821 and over	TOTAL
13002	Financial Managers													
15021	Mining, Quarrying, and Oil & Gas Well Drilling Managers													
19005	General Managers and Top Executives													
21114	Accountants and Auditors													
22105	Metallurgists and Metallurgical, Ceramic, and Materials Engineers													
22108	Mining Engineers, Including Mine Safety													
22311	Surveyors and Mapping Scientists													
22514	Drafters													
24111	Geologists, Geophysicists, and Oceanographers													
…														
98300	Helpers, Construction Trades and Extractive Workers													
98700	Freight, Stock, and Material Movers, Hand													

Table 1: Variance Decomposition

	(1)	(2)
R^2: W_{iej} = Occ Dummies	.2870	.5466
R^2: W_{iej} = Est Dummies	.4955	.4955
R^2: W_{iej} = Occ + Est	.6468	.7552
R^2: W_{iej} = Occ*Est	.7252	.8798
Occupation	.1513	.2597
Joint Occup & Estab	.1357	.2869
Establishment	.3598	.2086
Job Cell	.0784	.1246
Individual	.2748	.1202
One-Digit Occupation	Yes	
Five-Digit Occupation		Yes

Source: OES unweighted microdata. 34,453,430 individuals.
Wages are measured in natural logarithms: Mean=2.5133, Std.Dev.=0.5446.
There are 7 One-Digit Occupations, 824 Five-Digit Occupations, and 573,586 establishments.

Table 2: The Effect of Observable Establishment Characteristics on EWDs

	(1)	(2)	(3)	(4)	(5)	(6)
R^2: $W_{iej} = X$	.0833	.0243	.0727	.1294	.2955	.3469
R^2: $W_{iej} = Occ + X$	.5884	.5499	.5684	.5658	.6104	.6515
Establishment Effect	.2086	.2086	.2086	.2086	.2086	.2086
Explained	.0418	.0033	.0218	.0192	.0638	.1049
Unexplained	.1668	.2053	.1868	.1894	.1448	.1037
County Controls	Yes					Yes
Age Controls		Yes				Yes
Size Controls			Yes			Yes
Major Industry Controls				Yes		Yes
4-Digit Industry Controls					Yes	Yes

See notes to Table 1.
There are 3,194 counties, 5 age categories, 9 size categories, 10 major industries, and 937 4-digit industries.

Table 3a: Variance Decomposition, by Major Industry

	Agricult	Mining	Constr	Manuf	TCPU	Whole sale	Retail	FIRE	Services	Public Admin
R^2: W_{iej} = X	.2819	.4187	.2511	.3542	.3114	.1612	.1912	.2032	.2937	.2207
R^2: W_{iej} = Occ	.5960	.4858	.3332	.5112	.4496	.4778	.4575	.5319	.6075	.4282
R^2: W_{iej} = Occ + X	.6596	.7042	.5325	.6765	.5826	.5547	.5516	.6111	.6769	.5615
R^2: W_{iej} = Est	.4340	.5284	.4556	.5144	.4844	.3880	.3784	.3465	.4360	.2909
R^2: W_{iej} = Occ + Est	.7666	.7829	.7017	.7855	.7171	.7063	.6932	.7028	.7630	.6111
R^2: W_{iej} = Occ*Est	.8921	.9114	.8595	.9110	.8565	.8789	.8466	.8376	.8802	.7626
Occupation	.3326	.2545	.2461	.2711	.2327	.3183	.3148	.3563	.3270	.3202
Joint Occup & Estab	.2634	.2313	.0871	.2401	.2169	.1595	.1427	.1756	.2805	.1080
Establishment	.1706	.2971	.3685	.2743	.2675	.2285	.2357	.1709	.1555	.1829
Explained	.0636	.2184	.1993	.1653	.1330	.0769	.0941	.0792	.0694	.1333
Unexplained	.1070	.0787	.1692	.1090	.1345	.1516	.1416	.0917	.0861	.0496
Job Cell	.1255	.1285	.1578	.1255	.1394	.1726	.1534	.1348	.1172	.1515
Individual	.1079	.0886	.1405	.0890	.1435	.1211	.1534	.1624	.1198	.2374
# Individuals	268,958	180,110	1,358,346	6,020,917	1,895,225	1,568,727	4,367,477	1,553,429	10,914,875	6,325,366
# Establishments	10,995	3,744	47,434	73,390	31,136	53,433	134,886	36,408	167,371	14,789
# 5-Digit Occupations	229	287	391	643	502	559	534	409	759	669

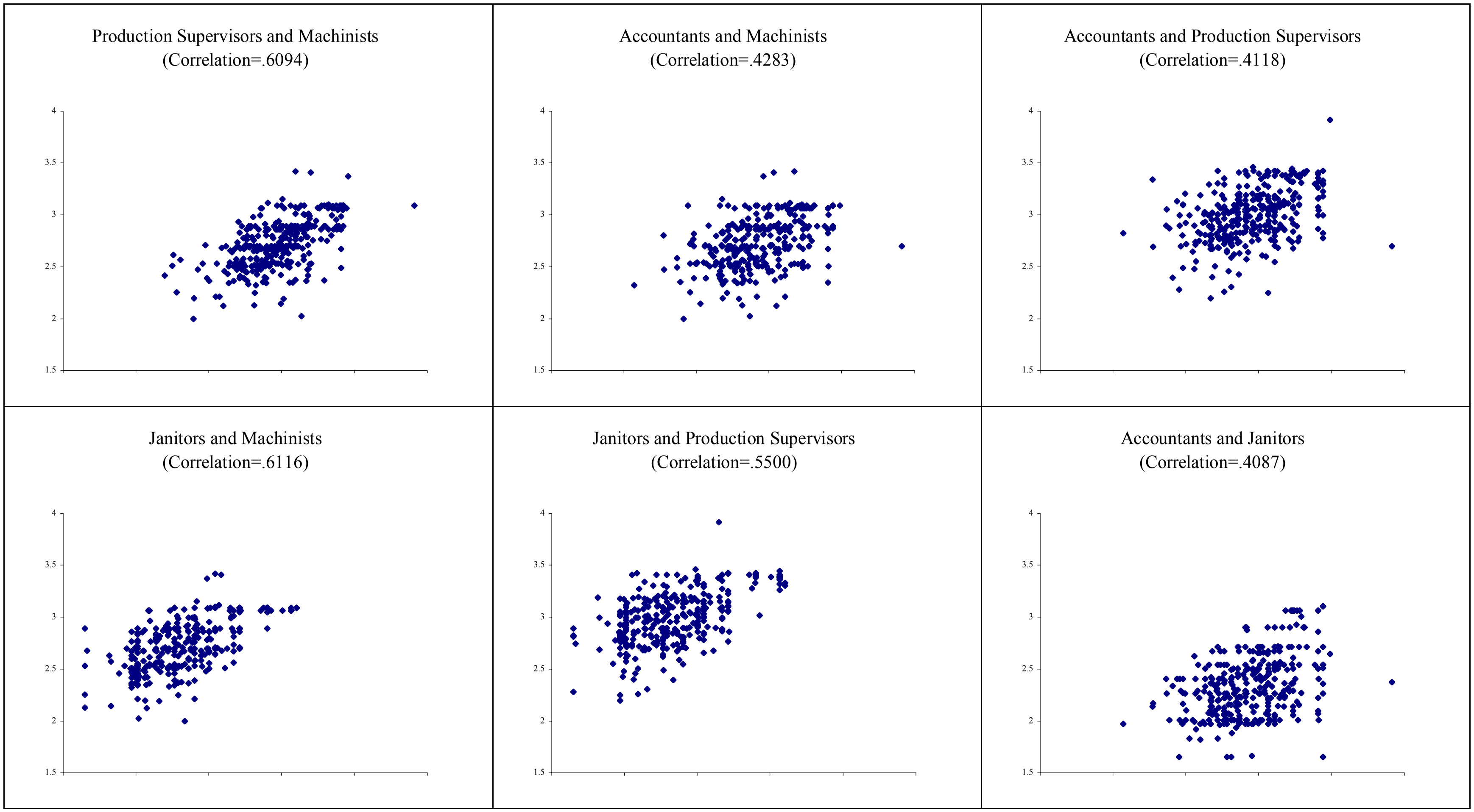
See notes to Tables 1 and 2.
Explanatory variables "X" are county, age, size, and 4-digit industry.

Table 3b: Variance Decomposition, by Establishment Size

	Size =1	Size 2-9	Size 10-15	Size 16-25	Size 26-50	Size 51-100	Size 101-250	Size 251-500	Size >500
R^2: $W_{iej} = X$	.6535	.2756	.2844	.3032	.3191	.3335	.3373	.3630	.3042
R^2: $W_{iej} = Occ$	.4735	.4692	.5082	.5366	.5575	.5666	.5670	.5826	.5438
R^2: $W_{iej} = Occ + X$	.8125	.5595	.5946	.6213	.6361	.6505	.6586	.6888	.6590
R^2: $W_{iej} = Est$	1.000	.5392	.4991	.4940	.4994	.5022	.4958	.4932	.3858
R^2: $W_{iej} = Occ + Est$	1.000	.7684	.7589	.7626	.7646	.7655	.7632	.7714	.7088
R^2: $W_{iej} = Occ*Est$	1.000	.9270	.9136	.9079	.9008	.8960	.8875	.8843	.8288
Occupation	.0000	.2292	.2598	.2686	.2652	.2633	.2674	.2782	.3230
Joint Occup & Estab	.4735	.2400	.2484	.2680	.2923	.3033	.2996	.3044	.2208
Establishment	.5265	.2992	.2507	.2260	.2071	.1989	.1962	.1888	.1650
Explained	.3390	.0903	.0864	.0847	.0786	.0839	.0916	.1062	.1152
Unexplained	.1875	.2089	.1643	.1413	.1285	.1150	.1046	.0826	.0498
Job Cell	.0000	.1586	.1547	.1453	.1362	.1305	.1243	.1129	.1200
Individual	.0000	.0730	.0864	.0921	.0992	.1040	.1125	.1157	.1712
# Individuals	3,149	1,098,076	1,292,496	1,806,070	3,073,260	3,890,886	5,477,999	3,880,169	13,931,325
# Establishments	3,149	177,200	106,272	90,111	86,388	55,087	36,111	11,280	7,988
# 5-Digit Occupations	377	791	802	806	815	819	821	812	816

See notes to Tables 1 and 2.
Explanatory variables "X" are county, age, size, and 4-digit industry.

Figure 2: Mean Occupational Wages, Manufacturing Industry



Source: OES unweighted microdata. Wages are measured in natural logarithms.
The axes in each chart are the average log wages of an occupation.
Sample is 338 establishments in the manufacturing industry with at least two employees in each of the following 5-digit occupations: Machinists, Production Supervisors, Accountants, and Janitors.

Table 4: Correlation of Mean One-Digit Occupational Wages Within Establishments

	Management	Professional	Sales	Clerical	Services	Agricultural	Production
Management	1	.5054	.5696	.4503	.3510	.3668	.3790
	1	.3964	.3668	.3346	.2041	.1798	.1935
	(N=378,960)	(N=190,508)	(N=177,866)	(N=309,002)	(N=123,393)	(N=29,415)	(N=234,127)
Professional		1	.4515	.4788	.4237	.3625	.4671
		1	.2249	.3604	.2900	.1293	.2315
		(N=242,710)	(N=95,201)	(N=212,116)	(N=91,243)	(N=20,786)	(N=126,181)
Sales			1	.5004	.3822	.3869	.5020
			1	.2072	.0912	.2273	.2469
			(N=263,965)	(N=179,827)	(N=67,313)	(N=12,940)	(N=145,992)
Clerical				1	.5138	.4904	.4878
				1	.4387	.3054	.3033
				(N=410,387)	(N=128,401)	(N=32,757)	(N=255,165)
Services					1	.5827	.4602
					1	.3351	.2591
					(N=173,193)	(N=17,470)	(N=88,471)
Agricultural						1	.5780
						1	.3447
						(N=41,203)	(N=25,329)
Production							1
							1
							(N=316,958)

Source: OES unweighted microdata. 573,586 establishments. Wages are measured in natural logarithms.
Upper Correlation: No Controls for Establishment Characteristics. Lower Correlation: Controls for County, Age, Size, and Major Industry.

Appendix Table 1: Number of OES Wage Intervals per Job Cell (using 5-digit occupational data)

		Size =1	Size 2-9	Size 10-15	Size 16-25	Size 26-50	Size 51-100	Size 101-250	Size 251-500	Size >500
1 Wage Interval	62.0%	100%	78.5%	72.5%	68.5%	63.7%	60.5%	56.1%	50.2%	37.2%
2 Wage Intervals	21.9%		16.6%	19.4%	20.9%	22.5%	23.0%	23.8%	24.9%	24.5%
3 Wage Intervals	9.0%		3.8%	5.6%	6.8%	8.3%	9.5%	11.0%	13.0%	16.5%
4 Wage Intervals	4.1%		0.9%	1.8%	2.5%	3.4%	4.1%	5.2%	6.7%	10.3%
5 Wage Intervals	1.8%		0.2%	0.5%	0.9%	1.3%	1.8%	2.4%	3.1%	5.9%
6-11 Wage Intervals	1.3%		0.0%	0.2%	0.4%	0.7%	1.1%	1.5%	2.1%	5.6%
# Establishments	573,586	3,149	177,200	106,272	90,111	86,388	55,087	36,111	11,280	7,988
# Occupations	824	377	791	802	806	815	819	821	812	816
# Est-Occ Job Cells	4,672,533	3,149	570,569	513,548	584,286	769,135	737,501	724,029	333,670	436,646
# Est-Occ-Wage Intervals	7,778,248	3,149	728,316	713,824	862,849	1,219,805	1,234,635	1,298,082	651,500	1,066,088
Average # Wage Intervals per Est-Occ Job Cell	1.66	1.00	1.28	1.39	1.48	1.59	1.67	1.79	1.95	2.44
R^2: W_{iej} = Occ*Est	.8798	1.000	.9270	.9136	.9079	.9008	.8960	.8875	.8843	.8288