The Role of Establishments (and the jobs therein) in Wage Inequality

Elizabeth Weber Handwerker
U.S. Bureau of Labor Statistics

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

PRELIMINARY AND INCOMPLETE

July 21, 2011

We are grateful to seminar participants at George Washington University, the Society of Government Economists meetings, the Society of Labor Economists meetings, the Berkeley Labor Lunch, and the Center for Economic Studies at the Census Bureau for comments on earlier versions of this work. All views expressed in this paper are those of the authors and do not necessarily reflect the views or policies of the U.S. Bureau of Labor Statistics.

I. Introduction

Income, earnings, and wage inequality have been growing for US workers since the late 1970s. This growth can be observed in a variety of measures – in total annual income (e.g. Piketty & Saez, 2003), in total compensation (e.g. Pierce, 2001, 2008), and in hourly wages (e.g. Katz and Autor, 1999 and Lemieux, 2006). While the largest rise in this inequality, particularly in the lower half of the earnings and income distributions, occurred during the 1980s, income and earnings have continued to grow less equal in the upper parts of the distribution in recent years.

An enormous literature has examined the composition and sources of this growing inequality, particularly in the 1980s, using data on individual workers and their characteristics. This work has addressed the changing composition of the workforce and changing returns to education and experience (Bound and Johnson, 1992, Katz and Murphy, 1992, Lemieux, 2006), and the growing inequality within education and skill groups (Juhn, Murphy, and Pierce, 1993, Katz and Autor, 1999). Growing inequality has been attributed to the differential impact of technology on differing portions of the worker skill distribution, referred to as 'Skill Biased Technology Change' (Juhn, Murphy, and Pierce, 1993, Acemoglu, 2002, Autor, Katz, and Kearney 2006, 2008), and to changing labor market institutions such as declining unionization levels (e.g. Lemieux, 2008), the declining real value of the minimum wage (e.g. Card and DiNardo, 2002, Lee, 1999), and the growing fraction of workers subject to performance-based pay from their employers (e.g. Lemieux, MacLeod, and Parent, 2009). Although these explanations for growing inequality are concerned with the policies and incentives faced by employers, this literature uses worker microdata containing little if any information on the businesses employing these workers.

A second smaller literature has used employer data to study growing wage inequality. This work builds on the evidence showing that establishments play an important role in determining individual wages (Groshen, 1991a, 1991b, Bronars and Famulari, 1997, Abowd, Kramarz, and Margolis, 1999, and Lane, Salmon, and Spletzer, 2007). Several authors have used employer microdata to study growing earnings variance in the U.S. from the mid-1970s to the early 2000s, and have found that the increasing variance is due more to variation between establishments than to variation within establishments (Davis and Haltiwanger, 1991, Dunne, Foster, Haltiwanger, and Troske, 2004, and Barth, Bryson, Davis, and Freeman, 2009). This literature has relied on combining measures of total variation in wages from worker microdata with measures of establishment mean wages from employer microdata, with limited information on the distribution of wages or worker characteristics within these establishments.

To these literatures we bring the Occupational Employment Statistics (OES) Survey microdata. The OES data is collected from a large annual survey of establishments, and contains information both on establishment characteristics and on the wage and occupational distributions of the employees within surveyed establishments. The OES data allow us to decompose increasing wage inequality in the U.S. into its within and between establishment components using a single source of wage information. We can also assess the impact of changing employer characteristics on the overall distribution of wages, and (in future work

not yet complete) we can assess the contribution of the changing distribution of employment by occupation within establishments on the overall wage distribution. In particular, we will be able to quantify the impact of 'contracting out' and other forms of establishment specialization on wage inequality in recent years.

We believe that employer-based explanations of increasing wage inequality warrant further investigation because much of the recent rise in inequality remains unexplained. Broad institutional forces such as the changing real value of the minimum wage or de-unionization may play a role in understanding the growth of inequality during the 1980s, but these do not appear to be as relevant for explaining the recent trends in inequality. In this paper, we examine the role that establishment characteristics play in explaining increasing wage inequality. In particular, our analysis (not yet completed) focuses on the role of the composition of employment by occupation within establishments. Whether rising inequality is driven by skill-biased technological change, by changes in labor market institutions, or by changes in employer-specific pay policies, our work can address the form of these changes. We can measure whether employers that pay divergent wages are employing workers in the same or in differing occupations. We can measure whether employers with more and less variation in wages among their employees have more and less variation in occupations among these employees. And we can also assess the impact of changes in occupational composition within establishments on the shape of the wage distribution.

We begin our analysis by showing that we can broadly replicate basic CPS wage inequality trends with the OES microdata. Using the OES as a unified source of wage data, we then confirm that almost all of the recent increase in wage inequality is between establishments rather than within establishments. We find that much of the rise in the between-establishment variance component is found in the upper tail of the mean establishment wage distribution, and we explore the industry, occupation, and size characteristics of establishments in this upper tail. We then show that we can explain up to 46% of the increase in private-sector wage inequality over the 1998-2007 time period by changes in the distribution of establishments by industry, establishment size, geography, and occupational concentration within establishments, and by changes in the national composition of occupational employment. 36% of the increase in private-sector wage inequality during this period can be explained by changes in the distribution of establishments in the fraction of their employment among the 3 aggregates of 'white-collar,' 'blue-collar,' and 'service' occupations alone. Clearly, the composition of employment within establishments by occupation matters for increasing wage inequality.

We are now in the process of extending the consistent series of OES data through 2010, adding the public sector to our analyses, and working to disaggregate the industry and occupational groups which we can examine throughout this time period. Our analysis of how the changing patterns of occupations within establishments has affected wage inequality, and what these patterns mean for various explanations of growing inequality, remains incomplete.

II. The OES and the CPS Data

Our empirical analysis relies on wage information supplied by establishments in the Occupational Employment Statistics (OES) survey. Since these data have been little used for studying wage inequality, we compare the OES wage data with wage information supplied by individuals in the more familiar Current Population Survey (CPS).[1]

IIa. The OES data

The Occupational Employment Statistics (OES) survey is designed to measure occupational employment and wages by geography and industry. Although the OES survey existed before 1996, wage information was collected beginning in 1996. Between 1996 and 2001, the survey was 400,000 establishments with a reference period of October, November, or December. Starting in November 2002, and continuing to the present, the survey is 200,000 establishments in May and 200,000 establishments in November. We have found that the May and November panels exhibit large amounts of seasonality in employment and wages, and thus this paper largely uses the Oct/Nov/Dec panels from 1996 to 2001 and the November panels from 2002 to 2007. Agriculture, private households, and the self-employed are not in scope for the OES survey. To date, we are excluding all government jobs (although we are working on adding these data to our analyses).

The OES survey form is a matrix with occupations on the rows and wage intervals in the columns. The first page of a sample OES survey form is given in Exhibit 1. For large establishments, the survey form lists between 50 to 225 detailed occupations (the occupations pre-printed on the survey form are based on the industry and the size of the establishment). Small establishments receive a blank survey form and write in descriptions of the work done by their employees. These employer-provided descriptions are then coded by state staff (as part of the OES Federal-State partnership) into occupations. The wage intervals on the OES survey form are given in both nominal hourly dollars and annual dollars, with the transformation between hourly and annual being 2080 hours per year. There were 11 wage intervals on the survey form in the years 1996 through 1998, and 12 wage intervals on the survey form in the years 1999 and later. The nominal wage intervals printed on the survey form changed in 1999 and again in 2005.

Establishments who complete the OES survey write in the number of employees in each occupation-wage cell of the matrix. For example, examining exhibit 1, an establishment may have three computer information systems managers – one earning $70,000, one earning $80,000, and one earning $90,000. For the row with this specific occupation, the establishment would write in “1” in the “H” column ($56,680 - $71,759) and “2” in the “I” column ($71,760 - $90,999). The OES program obtains the mean of each wage interval every year from the National Compensation Survey (NCS) at BLS. These mean wages are then assigned as the wage for the respective interval. Continuing with the example above, the

[1] Previous authors who have used the OES microdata for economic research are Abraham and Spletzer (2009a, 2009b), Dey, Houseman, and Polivka (2009), and Lane, Salmon, and Spletzer (2007).

wage assigned to the computer information system managers paid $80,000 and $90,000 (both in the "I" column) would be the same.

In preparing the OES data for analysis, the mean used in the microdata for the upper interval was $60 in 1996-1998, $70 in 1999-2001, and the actual mean wage from the NCS in 2002-2007. We multiplied the 1996-2001 upper interval values by 1.4 to correct for this topcoding. Approximately 25 percent of jobs in the OES data are imputed (with an establishment nonresponse rate of approximately 20 percent). In this initial version of the paper, we keep imputed data in our analysis (calculations suggest that our time series of wage variances is unaffected when we delete imputed establishments).

Much of the preparatory work for our analyses focuses on creating appropriate weights in order to examine the OES data in individual panels. The OES survey is designed to produce national-level estimates of employment based on combining 3 years of data for each estimate. Using the methodology described in Abraham and Spletzer (2009b), we reweight the data to November (or May) benchmarks of total employment by detailed industry and by broad industry and establishment size groups as reported in the Quarterly Census of Employment and Wages (QCEW).

We show the pdf of wages in the OES for 1999 and 2007 in Figure 1. Note that the 1999 and 2007 intervals differ both because the nominal wage intervals listed on the OES survey form change over time, and because the wage intervals are displayed here in (logged) real dollars.

IIb. Modifying the CPS data for comparison to the OES.

We follow Lemieux (2006, 2009) in using the CPS Outgoing Rotation Group (ORG) samples in our analysis of wage inequality. We define the hourly wage in the CPS as the reported hourly wage for those who are paid by the hour (approximately two-thirds of people), and as the ratio of usual weekly earnings divided by usual weekly hours for those not paid hourly. We implement Thomas Lemieux's usual hours imputation for those persons who report that their usual hours vary. Both hourly and weekly topcoded earnings are multiplied by 1.4. We trim all hourly earnings in the CPS by setting to missing all real hourly earnings less than $1 and greater than $100 (in November 2007 dollars). We also set to missing all imputed wage observations in the CPS. Figure 2 shows the effect of trimming and imputations on the variance of log real hourly wages in the CPS. Trimming the very few outliers in the hourly wage distribution has a very substantial effect on the time series of variance – this is apparent by comparing the solid lines to the dashed lines in Figure 2. Removing imputations slightly raises the variance in any given year, but the time trends of the variance of trimmed wages are similar whether imputed values are or are not included in the data.

We have tried to make the CPS and OES data as similar as possible. As stated earlier, jobs in agriculture, government, and private households are excluded from both data sources. Nevertheless, two significant differences remain. First, the CPS earnings data is based on an individual's main job (individuals with multiple jobs are only asked about their earnings on

their main job), whereas the OES earnings data is recorded for all jobs. Abraham and Spletzer (2009a) state that 3.9 percent of jobs in their CPS sample of jobs are second jobs. Second, the earnings questions in the CPS are not asked of the self-employed (either incorporated or unincorporated), whereas the OES includes earnings for the incorporated self-employed (essentially, business owners). Abraham and Spletzer (2009a) state that 4.4 percent of jobs in their CPS sample are incorporated self-employed jobs. We make no attempt in this paper to impute earnings for either second jobs or self-employed jobs in the CPS.

Three other differences between the CPS and OES data warrant mentioning. One difference is that the OES data is collected with a reference period of November (technically, Oct/Nov/Dec in 1998-2001, and either May or November in later years, although data collected in May is ignored in this paper), whereas the CPS data is collected from the ORG samples in all 12 months. A second difference is that the OES has an "intervalized" wage whereas the CPS has a "continuous" wage. We address this difference directly in the next subsection of this paper. Figure 3 shows that the two data sources have the same trend in log real hourly wages, although the OES has a slightly higher level of wages in all years, perhaps due to the inclusion of business owners in the OES but not in the CPS, or to our inclusion of sub-minimum wages in the CPS. A full comparison of topside employment from the two data sources, as well as a comparison of occupational and industrial employment, is given in Abraham and Spletzer, 2009b.

IIc. The Effects of Intervalized Wages for Analyzing the Variance of Wages

The wage data in the OES are collected as intervals. This means that the wage information in any given year is a series of 11 discrete values in 1996-1998 and 12 discrete values in 1999-2007 (wages assigned to the bottom intervals vary by state depending upon state-specific minimum wages).

We empirically examine the difference in variances when using a continuous wage versus using an interval wage by intervalizing the CPS continuous wage data. Specifically, we create an intervalized CPS wage as follows (this algorithm is described for the intervals shown in Exhibit 1, with appropriate modifications for other years):

- for all nominal hourly wages less than \$6.75 (appropriately trimmed by setting all values less than \$1 in real November 1998 dollars to missing), redefine these wages to be the mean wage computed from this subset of jobs;
- for all nominal hourly wages between \$6.75 to \$8.49…
  and so forth for 11 or 12 intervals to the last interval:
- for all nominal hourly wages greater than or equal to \$70.00 (appropriately trimmed by setting all values greater than \$250 in real November 1997 dollars to missing), redefine these wages to be the mean wage computed from this subset of jobs.

The variance for two log real hourly wage series from the CPS data is shown in Figure 4. The variance is computed for the continuous CPS wage and the intervalized CPS wage. The variance for the intervalized wage is, on average, 6.1 percent lower than the variance for the continuous wage. Approximately 80 percent of this difference is attributable to the lowest wage interval (and is, we speculate, the result of the curvature of the natural logarithm

function). We conclude from Figure 4 that although variance levels are lower when wages are reported in intervals, the time series properties of the variance of the intervalized wage variable mimic the time series properties of the continuous wage variable.

The overall distribution of employment by wage interval in the CPS (corresponding to the trimmed and non-imputed intervalized wage) in 1999 and 2007 is shown in Figure 5, along with the OES wage distributions in the same years. We observe that the intervalized CPS distributions look similar to the OES in each year.

III. Statistical Framework and Descriptive Statistics

IIIa. Basic Notation

We define w_i as the log real wage of observation "i" and define k_i as the weight for observation "i." While it is simplest to think of observation "i" as an individual, in the OES data "i" refers to an establishment-occupation-wage interval cell, where the weight k_i reflects the number of individuals that establishments report in the occupation-wage interval cell, multiplied by the sampling and benchmarking weights to make the set of establishments in the OES representative of all establishments in the United States.

We group these observations by characteristic(s) "j," which index of one or more establishment or employee characteristics. In the OES data, these characteristics include industry, size class, geography, establishment, occupation, and discrete combinations of these. As an example, let j be the industry of each establishment. Total employment E is the sum of the industry-specific employments E_j:

$$(1) \qquad E = \sum_i k_{ij} = \sum_j \sum_{i \in j} k_{ij} = \sum_j E_j \; .$$

Mean log real wage $\overline{w}$ is the weighted sum of the industry-specific mean log real wages $\overline{w}_j$:

$$(2) \qquad \overline{w} = \left(\frac{\sum_i k_{ij} w_{ij}}{\sum_i k_{ij}} \right) = \sum_j \left(\frac{E_j}{E} \right) \left(\frac{\sum_{i \in j} k_{ij} w_{ij}}{\sum_{i \in j} k_{ij}} \right) = \sum_j \left(\frac{E_j}{E} \right) \overline{w}_j \; .$$

The variance of log real wages V(w) can be written as:

$$(3) \qquad V(w) = \left(\frac{\sum_i k_{ij} \left(w_{ij} - \overline{w} \right)^2}{\sum_i k_{ij}} \right) = \left(\frac{\sum_i k_{ij} \left(w_{ij} - \overline{w}_j \right)^2}{\sum_i k_{ij}} \right) + \left(\frac{\sum_i k_{ij} \left(\overline{w}_j - \overline{w} \right)^2}{\sum_i k_{ij}} \right)$$
$$= \left[\sum_j \left(\frac{E_j}{E} \right) \left(\frac{\sum_{i \in j} k_{ij} \left(w_{ij} - \overline{w}_j \right)^2}{\sum_{i \in j} k_{ij}} \right) \right] + \left(\frac{\sum_j E_j \left(\overline{w}_j - \overline{w} \right)^2}{\sum_j E_j} \right)$$

$$= \left[\sum_j \left(\frac{E_j}{E} \right) V_j(w) \right] + V(\overline{w}_j) \,.$$

Equation (3) shows that the variance of log real wages in the total economy is the sum of the variance of log real wages within industries and the variance of log real wages between industries. The within industry variance is equal to the weighted sum of industry-specific variances, and the between industry variance is equal to the variance of weighted industry mean wages.

IIIb. Industry Level Descriptive Statistics

The key elements in the above equations are the industry-specific employments E_j, the industry-specific employment shares (E_j/E), the industry-specific average wages $\overline{w}_j$, and the industry-specific variances $V_j(w)$. Descriptive statistics from the OES data for these industry-specific terms are in Table 1. Table 1a has descriptive statistics for 1998, Table 1b has descriptive statistics for 2007, and Table 1c has descriptive statistics for the growth of these elements between 1998 and 2007. When defining industries, we use the same 13 industries as Abraham and Spletzer (2009b); these 13 industries are defined to be consistent between the SIC (used by OES from 1996 to 2001) and the NAICS (used by OES since 2002). Our choice of the years 1998 and 2007 correspond to the empirical work that follows in the next section.

In 1998, as shown in Table 1a, the five largest industry groups are manufacturing (16.4% of employment), retail trade (14.4% of employment), professional and business services (13.4% of employment), healthcare and social assistance (11.7% of employment), and wholesale trade, transportation, and utilities (11.2% of employment). Between 1998 and 2007 (see Table 1c), the share of employment in manufacturing fell by 4.7 percent, the share in wholesale trade, transportation, and utilities fell by 1.5 percent, the share in healthcare and social assistance increased by 1.9 percent, and the share in professional and business services rose by 2.6 percent. Between 1998 and 2007, total employment rose by 8.3 million from 104.8 to 113.1 million.

In 1998, the industry with the largest variance of log real wages is professional and business services (.4145, as shown in column 5 of Table 1a). The industry groups with the next largest variances are finance & real estate, and healthcare & social assistance. When we examine the industry contributions to total variance, professional and business services contributes the most to total variance in both 1998 and 2007 (17 percent in 1998 and 21 percent in 2007, as shown in the last column of Table 1a and Table 1b, respectively).

Looking back at equation (3), the total variance is the sum of the variance within industries and the variance between industries. In 1998, the variance within industries is .2911 and the variance between industries is .0516. The within industry component accounts for 85 percent of total variance in 1998 (.2911/.3427) and 83 percent of total variance in 2007 (.3148/.3792). In terms of growth of total variance between 1998 and 2007, the within industry component is growing by .0237 and the between industry component is growing by

.0128. The within industry variance accounts for 65 percent of the growth in total variance (.0237/.0365).

In the total economy, the variance of wages rose from .3432 in 1998 to .3793 in 2007. All industries in the OES data exhibit increasing wage variance between 1998 and 2007 – see column 5 of Table 1c. The largest increases in industry-specific variance are in educational services, rising from .2969 in 1998 to .3913 in 2007, and in arts and entertainment, rising from .2285 in 1998 to .3062 in 2007. However, educational services and arts and entertainment are two of the smallest industries as measured by employment shares (1.8 percent and 1.4 percent in 1998, respectively). Examining the large industries as measured by employment share, the largest increases in variance between 1998 and 2007 occurred in manufacturing, an increase of .0419, and in professional and business services, an increase of .0303. When examining the contribution to the growth in total variance (presented in column 8 of Table 1c), 53 percent of the growth in total variance is attributable to the professional and business services industry. The professional and business services industry has an increasing employment share, an increasing mean log real wage, and an increasing variance of log real wages. The next largest contributor to the growth in variance is the healthcare and social assistance industry (with a contribution of 22 percent).

IIIc. The composition of occupations within establishments

A unique contribution of the OES data is the information it contains regarding the composition of occupations within establishments. Although the OES used its own occupational classification system through 1998, and the Standard Occupational Classification system from 1999 thereafter, and substantial changes in the classification of employees into managerial and non-managerial occupations took place from 1999-2001, we can group occupations into 19 consistent occupational groups across the full 1998-2007 period. These groups are the "Professional" groups of Management; Business and Financial Operations; Engineering; Life, Physical, and Social Science; Computer and Math; Healthcare; and Other Professional and Technical occupations, the "Service" groups of Sales and Related; Office Administration; Protective Service; Food And Beverage; Cleaning and Building Services; and All Other Services, and the "Production" groups of Production Supervisors; Maintenence and Repair; Construction and Extraction; Production; Transportation and Moving; and Production Helpers. Neither these 19 consistent broad occupational groupings nor their aggregations into the 3 categories of 'white-collar,' 'blue-collar,' and 'services' occupations are standard occupational groupings used or endorsed by the OES program.

For each establishment, we can calculate a Herfindahl index of occupational agglomeration, using the formula $H = \sum_{1}^{k} \left(\frac{Occupation_k \ Employment}{Total \ Employment} \right)^2$, where k can be either the consistent broad groups of 19 occupations or the aggregations into 3 occupational groups.

We find that both of these Herfindahl indices are generally growing during the 1998-2007 period, indicating that establishments are generally becoming more occupationally uniform. As shown in Figures 5.1 and 5.2, this pattern is driven by larger establishments. Larger

establishments are becoming more occupationally uniform, while very small establishments (those with 1-4 employees) are becoming more occupationally diverse. We also find some variation in these patterns by industry. Manufacturing and retail establishments are employing a greater mix of occupations over this time period, while Business Services, Health, Arts, Education, Food, Information, and Mining establishments are employing a more uniform group of occupations. Other industries show an increase in Herfindahl3 but a decrease in Herfindahl19, or vice versa.

Overall, during the 1998-2007 period, Herfindahl3 measures grew more than Herfindahl19 measures, although Herfindahl3 measures can vary between 1/3 and 1 and Herfindahl19 measures can vary between 1/19 and 1. Establishments made greater changes in how they allocated their employment between the 'white-collar,' 'blue-collar,' and 'services' aggregates than between occupations within these aggregates. To better understand these changes, we classify establishments by a variable we call "Occupational Share". This variable takes on differing values for establishments that

1) Employ at least 33% of their employees in each of 'white-collar,' 'services,' and 'blue-collar' (WBS)
2) Employ at least 33% of their employees in 'white-collar' occupations, and less than 33% of their employees in other occupations (W)
3) Employ at least 33% of their employees in 'services' occupations, and less than 33% of their employees in other occupations (S)
4) Employ at least 33% of their employees in 'blue-collar' occupations, and less than 33% of their employees in other occupations (B)
5) Employ at least 33% of their employees in 'white-collar' occupations, at least 33% of their employees in 'services' occupations, and less than 33% of their employees in other occupations (WS)
6) Employ at least 33% of their employees in 'white-collar' occupations, at least 33% of their employees in 'blue-collar' occupations, and less than 33% of their employees in other occupations (WB)
7) Employ at least 33% of their employees in 'services' occupations, at least 33% of their employees in 'blue-collar' occupations, and less than 33% of their employees in other occupations (SB)

Using this crude measure, we find that over the 1998-2007 period, employment in WBS, B, and WS establishments fell, while employment in W, S, and WB establishments rose (employment in SB establishments was unchanged).

We also find a relationship between the wages earned within our 19 broad occupational groups, and that occupation's share of total employment within the establishment. For the occupational groups that are part of the 'services' and 'blue-collar' aggregates, wages are generally higher when these occupations represent less of the establishment's total employment, and these wage-occupational concentration gradients steepened between 1998 and 2007. Thus, in the next section, we study whether these measures of occupational concentration can explain some portion of the growth in overall wage variation between 1998 and 2007.

IV. Empirical Results regarding Wage Inequality

IVa. The Time Series of the Variance of Wages

We now examine the time series of the variance of log real hourly wages from the OES data. Figure 6 adds the OES to the two CPS series that were graphed in Figure 4. We find that the OES time series of wage variances is broadly similar to the CPS time series of wage variances. However, the OES time series has several outliers that are immediately visible and concerning.

We believe we understand some of the outliers that cause the lack of smoothness in the OES time series. The years 1996 and 1997 in Figure 6 concern us. 1996 was the first year wages were collected in the OES, and we should always be cautious of any initial collection. 1997 was the year that very small establishments were added (in 1996, establishments with less than 5 employees were not sampled, and the OES weights for the establishments with 5-9 employees were modified to account for this). Although we have weighted the OES data to the QCEW industry by size employment counts, our analysis of the data shows the 1997 outlier exists in most of the size classes but is most obvious in establishments with less than 4 employees. For these reasons, we do not use observations for 1996 and 1997 in the remainder of our work.

Also noticeable in Figure 6 is that the 2002 OES wage variance is a clear outlier in the time series. Many survey changes happened in 2002, including a move from surveying 400,000 establishments each fall to 200,000 establishments each May and November, and a switch from SIC to NAICS industry coding. Neither of these strike us as immediate explanations for the 2002 outlier in wage variance, particularly considering we have benchmarked OES employment to QCEW detailed industry by size employment. We note that usually high wage variance in 2002 occurs primarily in 6 industries: wholesale trade, construction, manufacturing, information, finance, and professional and business services. It also occurs primarily in establishments with more than 250 employees. We intend to study this further.

Dropping the 1996 and 1997 data, and linearly interpolating between 2001 and 2003, we get the time series of wage variances in Figure 7. In the top panel of Figure 7, we see that the time series of CPS and OES wage variances are similar between 1998 and 2004, and then diverge somewhat in 2005 and 2006.

IVb. The Effects of Changing Composition

In this subsection, we conduct a reweighting exercise in order to understand how much of the increasing wage inequality in the OES between 1998 and 2007 is due to changes in the employment composition of observable characteristics such as industry, establishment size, geography, and occupation. In addition to calculating a counterfactual variance that holds constant the employment shares of the observables at their initial values, we also calculate a counterfactual wage distribution based on the discrete OES wage intervals. This

allows us to determine which parts of the wage distribution are affected by the changing composition of employer characteristics and occupational mix.

An example may help understand what we expect to learn from this reweighting exercise. We know that there has been employment polarization during the last 10-20 years: see Autor, Katz, and Kearney (2006), Goos and Manning (2007), Goos, Manning, and Salomons (2009), and Abraham and Spletzer (2009b). Using the OES data, and defining "jobs" by industry and occupation, Abraham and Spletzer (2009b) show that the share of both low-wage and high-wage jobs has risen from 1996 to 2004, whereas the share of middle-wage jobs has fallen (employment growth has polarized). This polarization should lead to increasing wage inequality. Our reweighting exercise holds constant the employment composition of occupations and industries at their 1998 values when calculating the variance of log real hourly wages in 2007, and the resulting counterfactual wage variance quantifies the magnitude of polarized employment growth on the increasing wage variance. This example can be generalized by asking how much of the increasing wage variance between 1998 and 2007 can be explained by changes in the observable characteristics, and where in the wage distribution this explained increase in variance is coming from.

Recall from our earlier notation that k_{ij} is the weight for observation "i" in some aggregation "j," where "j" denotes industry, establishment size, geography, or occupation. We define p_j as the percent of total employment in aggregation "j". Adding a superscript to denote a given year, we have:

$$\text{(4a)} \qquad p_j^{year} = \left[\frac{\sum_{i \in j} k_{ij}^{year}}{\sum_i k_{ij}^{year}} \right] = \left[\frac{E_j^{year}}{E^{year}} \right].$$

To hold the employment shares of "j" constant at their 1998 values when calculating the wage variance in 2007, we reweight the 2007 data by defining a weight adjustment factor α_j as:

$$\text{(4b)} \qquad \alpha_j = \left(\frac{p_j^{1999}}{p_j^{2007}} \right) = \left[\frac{E_j^{1999} / E^{1999}}{E_j^{2007} / E^{2007}} \right].$$

With new weights of (α_j * k_{ij}) instead of k_{ij}, we recalculate the variance in equation (3) and call this our counterfactual variance.

The variance is only a single statistic that summarizes the wage distribution. We would like to know how the distribution is changing when we hold characteristics fixed at their 1998 values. Given the OES collection of wage information in intervals, this is equivalent to asking how the employment of the various wage intervals is changing when we hold the characteristics fixed at their 1998 values. To do this, we implement the method of DiNardo, Fortin, and Lemieux (1996), calculating new weights for counterfactual distributions by estimating probit regressions for the probability that observations with particular characteristics appear in 1998 rather than 2007.

Estimates of the 2007 wage variance and employment by wage interval from data reweighted to hold constant select OES characteristics at their 1998 proportions are shown in Table 2. Without any reweighting of characteristics, in the 2007 date, the variance of log real hourly wages is .3786, with 9.6 million jobs in the first wage interval (nominal hourly wage <$7.50), 16.9 million jobs in the second wage interval (nominal hourly wage between $7.50 and $9.49), …, and 1.7 million jobs in the twelfth wage interval (nominal hourly wage ≥$80).

There have been major changes in distribution of employment by establishment size between 1998 and 2007. In 1998, 13.9 percent of employment was in establishments with 1-9 employees; this rose to 14.4 percent in 2007. Between 1998 and 2007, the share of employment in establishments with 10-99 employees increased from 40.1 percent to 42.1 percent. In 1998, 19.6 percent of employment was in establishments with 500 or more employees; this fell to 17.1 percent in 2007. The statistics in Row (1) of Table 2 show that this changing size distribution has little if any effect on the variance of log hourly real wages – when holding employment shares by establishment size constant at their 1998 value, the variance of wages rises relative to the baseline. Adding establishment size to other variables (as, for example, in row (8)) never increases the total amount of wage variance increase explained.

We also know that the distribution of employment by state changed from 1998 to 2007. The OES data (which we have not benchmarked by geography) shows large employment gains over this time period in California and Texas, and noticeable employment losses in several large Midwestern states. This shift in employment from Midwestern to the Southern and Western states has only a small effect on measured total variance, as shown in row (3) of Table 2. However, the changing geographic distribution of employment does have an impact on the distribution of wages that adds to other characteristics in explaining the total wage variance growth over this period, as shown in rows (8), (11), and (12), particularly in explaining the growth of employment in the upper tail of the wage distribution.

Reweighting employment by occupation (not an establishment characteristic) to its 1998 proportions, shown in row (7) of Table 2, explains a greater amount of the growth in logged real wage variance (36%) than any other single variable, and explains almost the same amount of variance as the other available variables combined. Adding occupation to other variables, as in rows (8) and (12), always increases the total amount of wage variance increase explained. The variance growth due to the changing occupational composition of the workforce is largely concentrated in the upper tail of the wage distribution. This is consistent with Firpo, Fortin, and Lemieux (2011).

We have 3 measures of occupational specialization: Herfindahl indices calculated from 3 occupational aggregates, Herfindahl indices calculated from 19 broad occupational categories, and the Occupational Share variable described in Section IIIc. We find that the Herfindahl-19 variable along does not explain ANY of the growth in total wage variance between 1998 and 2007, as shown in row (2) of Table 2, while the Herfindahl-3 variable alone explains 7% of the growth in total wage variance, as shown in row (4). The smaller impact of the more detailed variable suggest to us that more detailed decompositions of occupation will not help us explain inequality growth. Both Herfindahl variables affect the

amount of employment in the lower tail of the wage distribution. The Occupational Share variable alone explains 29% of the growth in total wage variance, affecting both the lower and upper tails of the wage distribution. Combined, these three measures of occupational specialization have greater power, explaining 37% of the growth in wage inequality, as shown in row (10) of Table 2. When Occupation is added to them, these four variables explain 44% of the total growth in wage inequality.

When reweighting employment by broad industry groups to its 1998 proportions, the 2007 wage variance falls to .3738, as shown in row (6) of Table 2. This reduction in variance is due to decreased employment in the lower wage intervals (and the highest wage interval) and increased employment in the middle of the wage distribution. Industry group explains more of the growth in wage variance between 1998 and 2007 than any other single establishment characteristic. However, the reweightings that explain the greatest amount of growth in wage variance, such as those in rows (11) and (12), do not include industry as a reweighting variable. The impact of industry on the growth of wage variance is completely captured by the combined impact of changes in the occupation, state, and occupational specialization distributions of employment. We suspect this may be because our industry variable is too aggregated to capture some important impacts of changes in industry composition on wage variance, and will work to refine our measure of industry in future.

In combination, these observable characteristics of establishments and their workers can explain as much as 46 percent of increasing wage inequality in the OES data between 1998 and 2007, as shown in row (12) of Table 2. In the next section of our paper, we examine the role of establishment effects in wage inequality more generally.

V. The Role of Establishments in Increasing Wage Inequality

In order for the establishment to have a role in explaining increasing wage inequality, we need to ask the basic question of whether an individual's wage is related to the establishment where he or she works. Although a strict interpretation of the competitive model would suggest no, there is a large literature which shows that establishment wage differentials exist, and are quite large – see Davis and Haltiwanger (1991), Groshen (1991a), Bronars and Famulari (1997), Abowd, Kramarz, and Margolis (1999), and Lane, Salmon, and Spletzer (2007). Groshen (1991b) lists five explanations for why wages might vary between employers: sorting, compensating differentials, random variations, efficiency wages, and rent sharing.

Va. The Within- and Between-Establishment Variance Decomposition

Our decomposition of log real hourly wages into its within- and between-establishment components is based on equation (3) above, where "j" is now defined as the indicator for the establishment. The empirical decomposition of the variance of wages into its within-establishment and between-establishment variance is given in Table 3 and Figure 8. The statistics in Table 3 and Figure 8 are averages across all years 1998-2007.

Examining the variance of wages in the total economy (the bottom row of Table 3), we see that in the average year, 45.7 percent of the variance is within establishments and 54.3 percent is between establishments. These are roughly comparable to other results in the literature that have used a single dataset: Bronars and Famulari (1997), using data from a supplement to the 1989 and 1990 White Collar Pay survey, found that 45 percent of variance is between establishments, and Lane, Salmon, and Spletzer (2007), using data from the 1996 and 1997 OES, found that 50 percent of variance is between establishments. Barth, Bryson, Davis, and Freeman (2009) use individual data from the 1977-2002 CPS and establishment data from the 1977-2002 Census Bureau's Longitudinal Business Database (LBD), and find that 55-70 percent of the variance in log earnings is between establishments.

Examining the results for specific industries in Table 3, we see that manufacturing has 54.5 percent of wage variance within establishments and 45.5 percent between establishments. Although our time period, source data, and methodology differ, this result echoes the previous literature. Davis and Haltiwanger (1991), using the 1975-1986 LRD in combination with the CPS, found that 50 to 58 percent of wage variance in manufacturing is between plants. Updating this result, Dunne, Foster, Haltiwanger, and Troske (2004) use the 1977-1992 LRD in combination with the CPS and find that 53 to 69 percent of wage variance in manufacturing is between establishments. And using data from the 1977-2002 LBD in combination with the CPS, Barth, Bryson, Davis, and Freeman (2009) find that on average 62 percent of variance in manufacturing is between establishments.

We are interested in the growth of the within- and between- establishment components of wage variance. Figure 9 presents the time series of within and between establishment variance for the total economy. The within-establishment variance is essentially constant over the years 1998 through 2007, with a low value of .1637 in 1998 and a high value of .1717 in 2003. The between-establishment variance shows a definite upward trend, starting at .1795 in 1998 and rising to .2114 in 2007. Interestingly, the 2002 outlier in total variance appears to be entirely attributable to the between-establishment component, which suggests that the sample of establishments in November 2002 is "different" than in other years.

Examining the growth from 1998 to 2007, the statistics underlying Figure 9 show that the within establishment variance rises by .0042 and the between establishment variance rises by .0319. These two statistics, along with the industry specific statistics, are shown in Table 4 and Figure 10. This finding that most of the growth in wage variance is between establishments is in accordance with the small but growing literature. Looking first at the total economy, Barth, Bryson, Davis, and Freeman (2009) find that the growth in the between-establishment variance is at least as large as the growth in overall wage dispersion between individuals. And combining individual-level wage data with company-level data from the UK, Faggio, Salvanes, and Van Reenan (2007) find that the between-establishment component of wages accounts for just about all of the increase in the overall variance growth.

The industry specific statistics in Table 4 and Figure 10 show that 10 of our 13 industries have growth in wage variance that is larger between establishments than within establishments (the exceptions are mining, retail trade, and educational services). With regard to manufacturing, which has been the focus of the earlier published literature, we find that 69

percent of the growth in variance is between establishments. Davis and Haltiwanger (1991) find that 48 percent of variance growth in manufacturing is between establishments, Dunne, Foster, Haltiwanger, and Troske (2004) find that 90 percent of variance growth in manufacturing is between establishments, and Barth, Bryson, Davis, and Freeman (2009) find that 27 percent (.034/.125 in Table 2) of variance growth in manufacturing is between establishments.

Two particularly interesting industries are Finance and Health. These industries have negative growth in within-establishment variance and positive growth in between establishment variance. These results suggest that the establishments in these industries are specializing into high and low wage establishments. If this is so, we would expect the occupational mix in establishments in these industries to have changed substantially between 1998 and 2007. We will test for such changes in a future version of this paper.

Vb. The Distribution of the Between-Establishment Variance

We have found that almost all of the increasing wage variance between 1998 and 2007 is due to the growth in between establishment variance rather than growth in within establishment variance. In this subsection, we examine the growth in the between-establishment variance component in more detail.

Recall from equation (3) that the between-establishment variance is the weighted variance of the establishment mean wage $\overline{w}_j$, where the weights are the size of the establishment E_j. We can further rewrite the between-establishment variance as the sum of the contribution to variance by establishments “j” according to the deciles “p” of the establishment mean wage distribution:

$$(6) \quad \left(\frac{\sum_j E_j \left(\overline{w}_j - \overline{w} \right)^2}{\sum_j E_j} \right) = \sum_{p=1}^{10} \sum_{j \in p} \left(\frac{E_j}{E} \right) \left(\overline{w}_j - \overline{w} \right)^2 .$$

The empirical quantification of equation (6) is given in Table 5. We report the first, fifth, ninth, and tenth deciles, and further break the first and tenth deciles into halves.

The first row of Table 5 shows the contribution to variance by establishments with mean log real hourly wages in the first five percent of the weighted establishment mean-wage distribution. In 1998, the very low-wage establishments contributed .0230 to the total variance of .1795, and in 2007, the very low-wage establishments contributed .0287 to the total variance of .2114. The growth of the variance contribution from the low wage establishments is .0057, or 17.9 percent of the total growth in variance.

The statistics in Table 5 show that 24.5 percent of the growth in between-establishment variance between 1998 and 2004 is attributable to establishments that pay in the bottom 10 percent of the establishment mean wage distribution, 31.6% of the growth is attributable to establishments between the 10th and the 90th percentiles, and 42.2 percent of the

growth is attributable to establishments that pay in the top 10 percent of the establishment mean wage distribution. Breaking this down somewhat further, we see that 31.7 percent of the growth in the between-establishment wage distribution is attributable to the very highest paying establishments – those that have average wages in the top five percent of the establishment mean wage distribution.[2]

In Table 6, we show the industry and size class distribution for the full OES sample in 2007 and for the sample of establishments in the upper five percent of the establishment mean wage distribution. The most noticeable industry statistic is that 40.7 percent of the highest mean-wage establishments are in professional and business services; this is in contrast to the 18.1 percent statistic in the full sample. With regard to the size class statistics in Table 6, it is not surprising that the largest establishments (100 or more employees) are more likely to be in the sample of high mean-wage establishments, since we know that large establishments have higher average wages. The most noticeable size class statistic is that 36.2 percent of the highest mean-wage establishments have 1-4 employees; this is much higher than their 22.1% representation in the full sample.

There is something unique about these very small establishments and their propensity to be in the sample of the highest mean-wage establishments. We believe that these establishments are likely specializing in high-skill tasks, and we will formally analyze this in a future version of this paper. We have looked at establishments in the professional and business services industry which also have 1-4 employees. In the full 2007 sample, these establishments have 30 percent of their employees working in the "Office and Administrative" support occupation (such as secretaries, bookkeepers, clerks, …). However, establishments in the top five percent of the establishment mean wage distribution have only 7.7 percent of their employees in this occupation.

VI. Explaining the Increase in Between Establishment Variance

In Section V, we showed that about 90% of the growth in total log real wage variance during the 1998-2007 period is growth in the variation of wages between establishments, particularly due to wage growth in the establishments with high mean wages. In Section IV, we showed that changes in the broad industry mix and geographic composition of establishments can explain about 17% of the total variance growth. We also showed that changes in the occupational composition of establishments (not of individual workers), can explain 37% of the total variance growth.

Several facts suggest a redistribution of establishments into narrow specialties within the broad industries we measure. In the Finance and Health industries, within-establishment variance is falling while between-establishment variance is rising. The Professional and

---

[2] This finding that approximately one-third of the growth of the across-establishment variance component is attributable to establishments with the highest average wages echoes Piketty and Saez (2003) and Saez, 2009. They find that the income share of the top decile has been increasing since the late 1970s. Although most of this increase is in the top percentile, composed of capital income, the 95^{th}-99^{th} percentile increasedas shares of total income during much of the 1980s and 1990s (See figure 2 of Saez, 2009). Incomes in the 95^{th}-99^{th} percentiles were composed mostly of wage income.

Business Services industry has a high concentration of establishments among those paying top 5% mean wages, although this industry does not have particularly high wages overall. The reweightings by 1998 characteristics that explain the greatest share of total variance growth do not include broad industry as one of the explaining characteristics. We can investigate the contribution of the changing distribution of establishments by much narrower industry groups to rising variance, but have not yet done so. Beginning with the 2002 OES survey, establishments were classified by 6 digit NAICS, and the OES staff converted much of the previous years' samples from SIC to 6 digit NAICS codes as well. We are working to bring these data into our analyses.

More generally, our results give circumstantial evidence consistent with both explanations of growing wage inequality related to technological changes and explanations related to institutional factors. Our finding that 37% of the total growth in the variance of logged real wages can be explained by the increasing occupational specialization of establishments is suggestive of a role for technological change, which may affect the structure of establishments. At the same time, although changes in the distribution of establishments by state explain only 5% of the total growth in overall wage inequality, adding state as an explanatory variable always increases the proportion of variance growth explained in our reweighting exercises. This suggests that institutional differences between states, such as state laws encouraging or discouraging union formation, may also play a small but measurable role in wage inequality growth.

In future work, we intend to bring in OES data from the public sector, and examine whether any change in the composition of workers between sectors is related to wage inequality growth. We intend to disggregate the OES data by industry in much greater detail, and examine the impact of changes in the distribution of more detailed industry on wage inequality. We intend to study whether employers paying differing wages within the same industries show any differences in the occupations of workers they employ. Finally, we have not yet taken any advantage of the panel structure of the OES data. We have noted that occupational specialization is growing most among the largest establishments. These largest establishments are also the most likely to be sampled repeatedly by the OES (as often as once every three years). We may be able to trace some of these large establishments over time, observing the extent to which changes in their wage and occupational structures happen in parallel.

REFERENCES

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

Abraham, Katharine G. and James R. Spletzer (2009a). "New Evidence on the Returns to Job Skills." *American Economic Review Papers and Proceedings*, Vol. 99, No. 2, pp. 52-57.

Abraham, Katharine G. and James R. Spletzer (2009b). "Are the New Jobs Good Jobs?" Forthcoming, in *Labor in the New Economy*, edited by Katharine G. Abraham, James R. Spletzer, and Michael Harper. University of Chicago Press.

Acemoglu, Daron (2002). "Technical Change, Inequality, and the Labor Market." *Journal of Economic Literature*, Vol. 40, No. 1, pp. 7-72.

Autor, David H., Lawrence F. Katz, and Melissa S. Kearney (2006). "The Polarization of the U.S. Labor Market," *American Economic Association Annual Meeting Papers and Proceedings*, pp. 189-194.

Autor, David H., Lawrence F. Katz, and Melissa S. Kearney (2008). "Trends in U.S. Wage Inequality: Re-Assessing the Revisionists." *Review of Economics and Statistics*, Vol. 90, No. 2, pp. 300-323.

Barth, Erling, Alex Bryson, James C. Davis, and Richard Freeman (2009). "The Dynamics of Workplace Wage Inequality in the United States, 1977-2002." Unpublished paper, http://online.ceu.hu/labor/cost/Papers/Bryson.pdf (accessed January 21, 2010).

Bound, John and George Johnson (1992). "Changes in the Structure of Wages in the 1980's: An Evaluation of Alternative Explanations." *The American Economic Review*, Vol. 82, No. 3, pp. 371-392.

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

Card, David and John E. DiNardo (2002). "Skill-Biased Technological Change and Rising Wage Inequality: Some Problems and Puzzles." *Journal of Labor Economics*, Vol. 20, No. 4, pp. 733-783.

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

Dey, Matthew, Susan Houseman, and Anne Polivka (2009). "What Do We Know About Contracting Out in the United States? Evidence from Household and Establishment Surveys." Forthcoming, in *Labor in the New Economy*, edited by Katharine G. Abraham, James R. Spletzer, and Michael Harper. University of Chicago Press.

DiNardo, John, Nicole M. Fortin, and Thomas Lemieux (1996). "Labor Market Institutions and the Distribution of Wages, 1973-1992: A Semiparametric Approach." *Econometrica*, Vol. 64, No. 5, pp. 1001-1044.

Dunne, Timothy, Lucia Foster, John Haltiwanger, and Kenneth R. Troske (2004). "Wage and Productivity Dispersion in United States Manufacturing: The Role of Computer Investment." *Journal of Labor Economics*, Vol. 22, No. 2, pp. 397-429.

Faggio, Giulia, Kjell G. Salvanes, and John Van Reenen (2007). "The Evolution of Inequality in Productivity and Wages: Panel Data Evidence." Unpublished paper, http://cep.lse.ac.uk/pubs/download/dp0821.pdf (accessed January 21, 2010).

Firpo, Sergio, Nicole M. Fortin, and Thomas Lemieux (2011). "Occupational Tasks and Changes in the Wage Structure." Working paper, http://faculty.arts.ubc.ca/nfortin/FFL_Occupations_feb2011.pdf

Goos, Maarten and Alan Manning (2007). "Lousy and Lovely Jobs: The Rising Polarization of Work in Britain," *Review of Economics and Statistics*, Vol. 89, No. 1, pp. 118-133.

Goos, Maarten, Alan Manning and Anna Salomons (2009). "Job Polarization in Europe," *American Economic Review Papers and Proceedings*, Vol. 99, No. 2, pp. 58-63.

Groshen, Erica L. (1991a). "Sources of Intra-Industry Wage Dispersion: How Much do Employers Matter?" *The Quarterly Journal of Economics*, Vol. 106, No. 3, pp. 869-884.

Groshen, Erica L. (1991b). "Five Reasons Why Wages Vary Among Employers." *Industrial Relations*, Vol. 30, No. 3, pp. 350-381.

Juhn, Chinhui, Kevin M. Murphy, Brooks Pierce (1993). "Wage Inequality and the Rise in Returns to Skill Wage Inequality and the Rise in Returns to Skill." *The Journal of Political Economy*, Vol. 101, No. 3, pp. 410-442.

Katz, Lawrence F. and Autor, David H. (1999). "Changes in the Wage Structure and Earnings Inequality." In *Handbook of Labor Economics* (volume 3A), edited by Orley Ashenfelter and David Card, Elsevier, pp. 1463-1555.

Katz, Lawrence F. and Kevin M. Murphy (1992). "Changes in Relative Wages, 1963-1987: Supply and Demand Factors." *The Quarterly Journal of Economics*, Vol. 107, No. 1, pp. 35-78.

Lane, Julia I., Laurie A. Salmon, and James R. Spletzer (2007). "Establishment Wage Differentials." *Monthly Labor Review*, Vol. 130, No. 4, April, pp. 3-17.

Lee, David (1999). "Wage Inequality in the United States during the 1980s: Rising Dispersion or Falling Minimum Wage?" *The Quarterly Journal of Economics*, Vol. 114, No. 3, pp. 977-1023.

Lemieux, Thomas (2006). "Increasing Residual Wage Inequality: Composition Effects, Noisy Data, or Rising Demand for Skill?" *American Economic Review*, Vol. 96, No. 3, pp. 461-498.

Lemieux, Thomas (2008). "The Changing Nature of Wage Inequality." *Journal of Population Economics*, Vol. 21, No. 1, pp. 21-48.

Lemieux, Thomas (2009). "What Do We Really Know About Changes in Wage Inequality?" Forthcoming, in *Labor in the New Economy*, edited by Katharine G. Abraham, James R. Spletzer, and Michael Harper. University of Chicago Press.

Lemieux, Thomas, W. Bentley MacLeod, and Daniel Parent (2009). "Performance Pay and Wage Inequality." *Quarterly Journal of Economics*, Vol. 124, No. 1, pp. 1-49.

Pierce, Brooks (2001). "Compensation Inequality." *Quarterly Journal of Economics*, Vol. 116, No. 4, pp. 1493-1525.

Pierce, Brooks (2008). "Recent Trends in Compensation Inequality." Forthcoming, in *Labor in the New Economy*, edited by Katharine G. Abraham, James R. Spletzer, and Michael Harper. University of Chicago Press.

Piketty, Thomas and Emmanuel Saez (2003). "Income Inequality in the United States, 1913-1998." *The Quarterly Journal of Economics*, Vol. 118, No. 1, pp. 1-39.

Saez, Emmanuel (2009). "Striking it Richer: The Evolution of Top Incomes in the United States (Update with 2007 estimates)." Unpublished paper, http://elsa.berkeley.edu/~saez/saez-UStopincomes-2007.pdf (accessed January 25, 2010).

Exhibit 1: Sample OES Survey Form

OCCUPATIONAL TITLE AND DESCRIPTION OF DUTIES		NUMBER OF EMPLOYEES IN SELECTED WAGE RANGES (Report Part-time Workers According to an Hourly Rate)												
		A	B	C	D	E	F	G	H	I	J	K	L	T
	Hourly (part-time or full-time)	under $6.75	$6.75 - 8.49	$8.50 - 10.74	$10.75 - 13.49	$13.50 - 16.99	$17.00 - 21.49	$21.50 - 27.24	$27.25 - 34.49	$34.50 - 43.74	$43.75 - 55.49	$55.50 - 69.99	$70.00 and over	Total Employment
	Annual (full-time only)	under $14,040	$14,040 - 17,679	$17,680 - 22,359	$22,360 - 28,079	$28,080 - 35,359	$35,360 - 44,719	$44,720 - 56,679	$56,680 - 71,759	$71,760 - 90,999	$91,000 - 115,439	$115,440 - 145,599	$145,600 and over	

Management Occupations

(Managers in this section have other managers/supervisors reporting to them.)

Occupation	Code	A	B	C	D	E	F	G	H	I	J	K	L	T
Chief Executives - Determine and formulate policies and provide the overall direction of companies or private and public sector organizations within the guidelines set up by a board of directors or similar governing body.	11-1011													
General and Operations Managers - Plan, direct, or coordinate the operations of companies or public and private sector organizations. Duties include formulating policies, managing daily operations, and planning the use of materials and human resources, but are too diverse in nature to be classified in any one functional area of management or administration.	11-1021													
Marketing Managers - Determine the demand for products and services offered by a firm and its competitors and identify potential customers. Develop pricing strategies with the goal of maximizing the firm's profits or share of the market.	11-2021													
Computer and Information Systems Managers - Plan, direct, or coordinate activities in such fields as electronic data processing, information systems, systems analysis, and computer programming.	11-3021													

Figure 1: Distribution of Log Real Hourly Wages, OES, 1999 and 2007

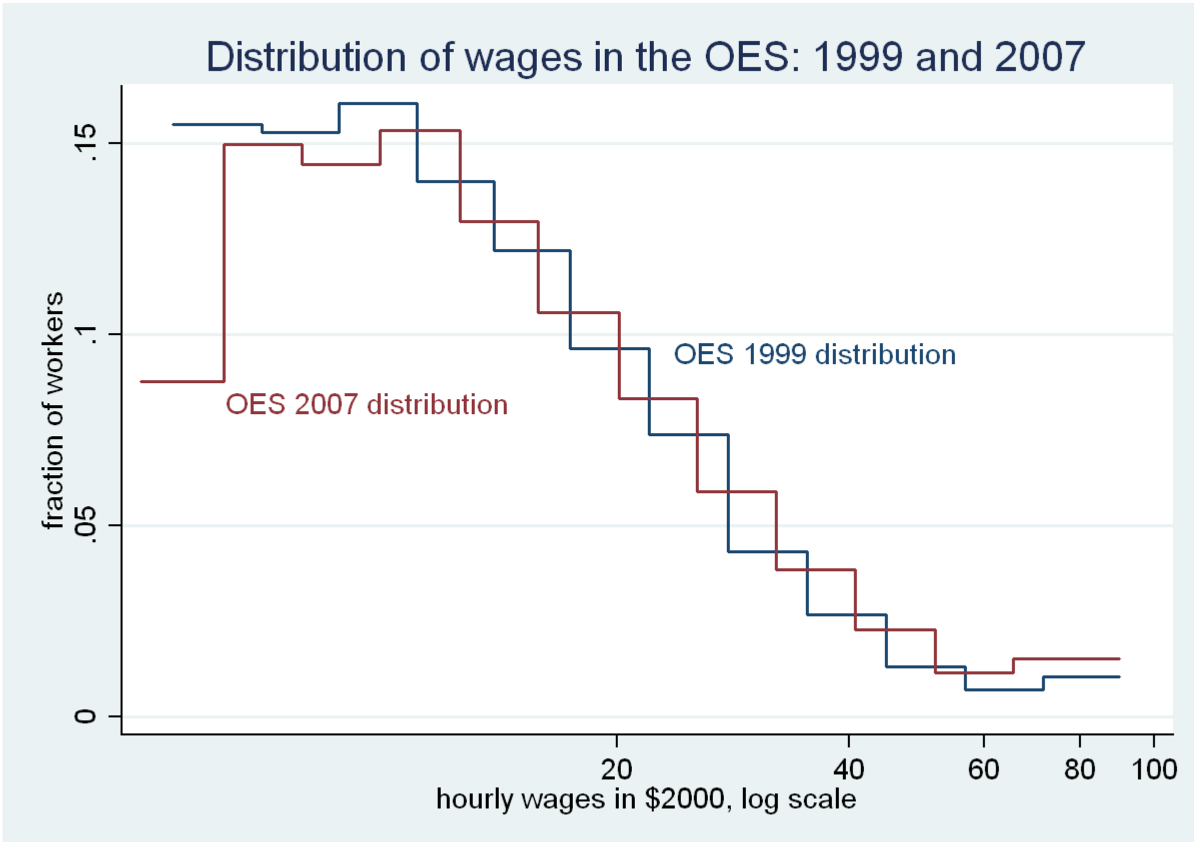


Figure 2: Variance of Log Real Hourly Wage, CPS, 1996-2007

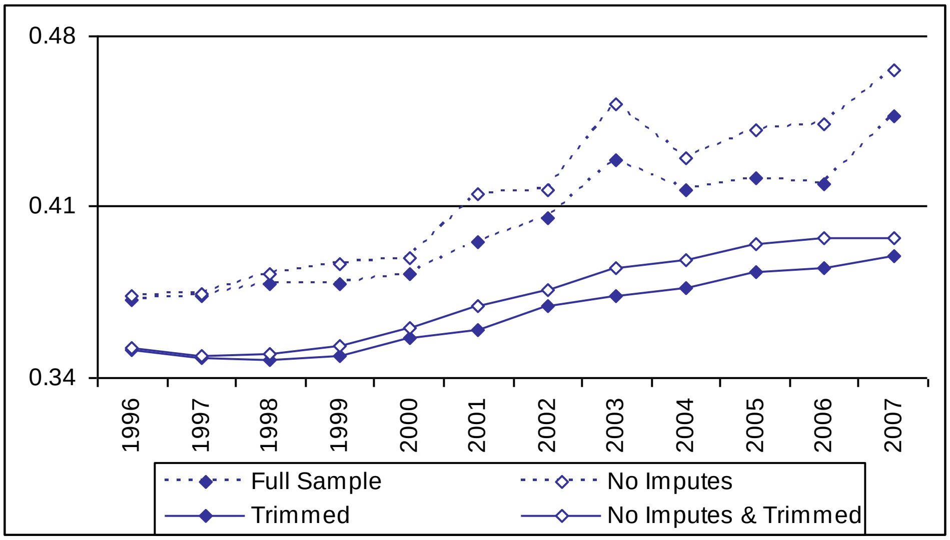


Figure 3: Mean Log Real Hourly Wage, CPS and OES, 1996-2007

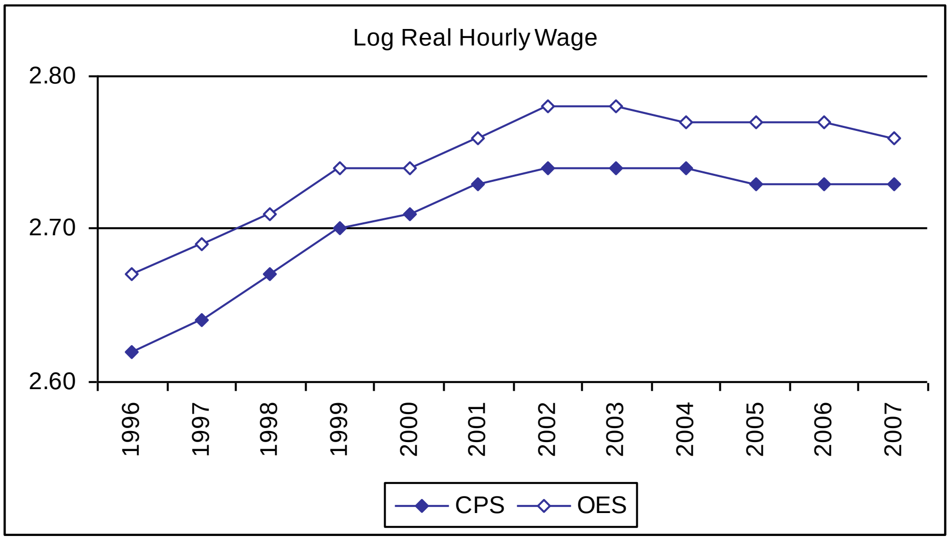


Figure 4: Variance of Log Real Hourly Wage, CPS, 1996-2007

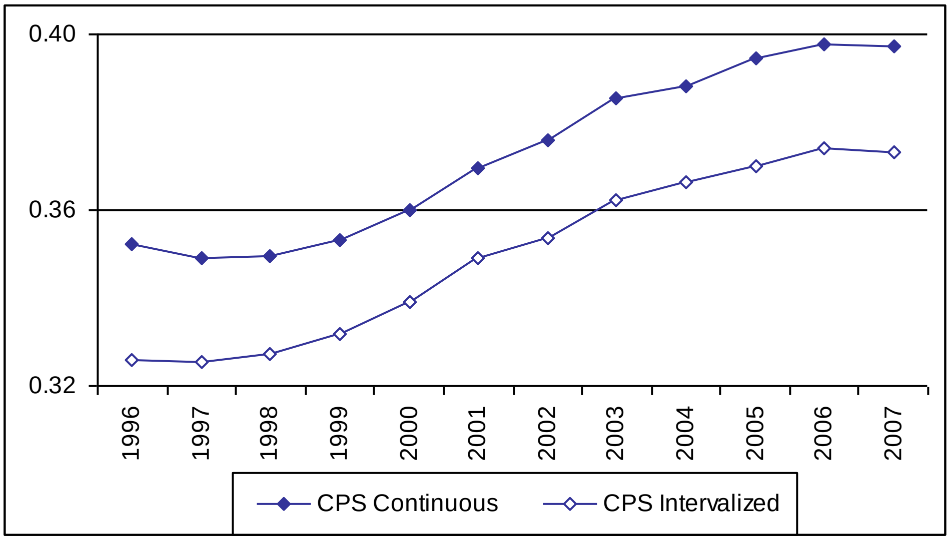


Figure 5: Distribution of Log Real Hourly Wages, OES and CPS, 1999 and 2007

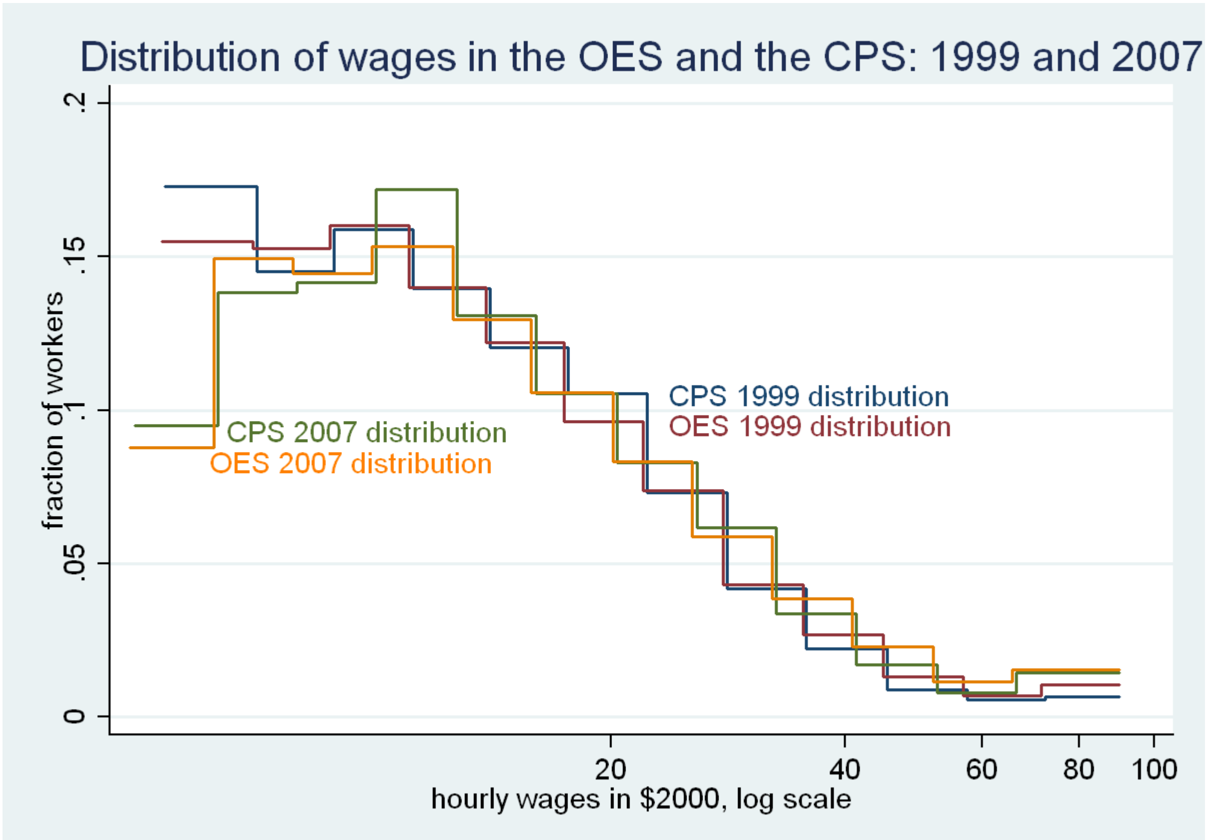


Table 1a: Descriptive Statistics, OES, By Industry, 1998

Industry j	(1) E_j (*1000)	(2) (E_j/E)	(3) $\bar{w}_j$	(4) $(E_j/E)^*$ $\bar{w}_j$	(5) $V_j(w)$	(6) $(E_j/E)^*$ $V_j(w)$	(7) $(E_j/E)^*$ $\left(w_j - \bar{w}\right)$	(8) Contribution to Total Variance
Mining	570	0.005	3.06	0.0153	0.2946	0.0015	0.0006	0.01
Retail Trade	15,135	0.144	2.43	0.3499	0.2226	0.0321	0.0113	0.13
Wholesale, Trans, Util	11,706	0.112	2.88	0.3226	0.3098	0.0347	0.0032	0.11
Construction	6,184	0.059	2.89	0.1705	0.2416	0.0143	0.0019	0.05
Manufacturing	17,159	0.164	2.84	0.4658	0.2689	0.0441	0.0028	0.14
Information	3,672	0.035	2.93	0.1026	0.3294	0.0115	0.0017	0.04
Finance & Real Estate	7,680	0.073	2.89	0.2110	0.3758	0.0274	0.0024	0.09
Prof & Bus Services	14,047	0.134	2.80	0.3752	0.4145	0.0555	0.0011	0.17
Educ Services	1,884	0.018	2.86	0.0515	0.2969	0.0053	0.0004	0.02
Health & Social Assist	12,259	0.117	2.77	0.3241	0.3608	0.0422	0.0004	0.12
Arts & Entertainment	1,471	0.014	2.41	0.0337	0.2285	0.0032	0.0013	0.01
Food & Lodging	9,500	0.091	2.20	0.2002	0.1133	0.0103	0.0237	0.10
Other Services	3,539	0.034	2.55	0.0867	0.2625	0.0089	0.0009	0.03
Sum	104,806	1.000		2.71		0.2911	0.0516	1.00

Table 1b: Descriptive Statistics, OES, By Industry, 2007

Industry j	(1) E_j (*1000)	(2) (E_j/E)	(3) $\bar{w}_j$	(4) $(E_j/E)^*$ $\bar{w}_j$	(5) $V_j(w)$	(6) $(E_j/E)^*$ $V_j(w)$	(7) $(E_j/E)^*$ $(\bar{w}_j - \bar{w})^2$	(8) Contribution to Total Variance
Mining	676	0.006	3.04	0.0182	0.2985	0.0018	0.0005	0.01
Retail Trade	16,039	0.142	2.44	0.3465	0.2320	0.0329	0.0145	0.13
Wholesale, Trans, Util	10,932	0.097	2.92	0.2832	0.3216	0.0312	0.0025	0.09
Construction	7,583	0.067	2.94	0.1970	0.2474	0.0166	0.0022	0.05
Manufacturing	13,193	0.117	2.88	0.3370	0.3108	0.0364	0.0017	0.10
Information	3,652	0.032	3.08	0.0986	0.3706	0.0119	0.0033	0.04
Finance & Real Estate	8,092	0.072	2.97	0.2138	0.3794	0.0273	0.0032	0.08
Prof & Bus Services	18,078	0.160	2.93	0.4688	0.4448	0.0712	0.0046	0.20
Educ Services	2,396	0.021	3.02	0.0634	0.3913	0.0082	0.0014	0.03
Health & Social Assist	15,356	0.136	2.83	0.3849	0.3682	0.0501	0.0007	0.13
Arts & Entertainment	1,836	0.016	2.52	0.0403	0.3062	0.0049	0.0009	0.02
Food & Lodging	11,341	0.100	2.23	0.2230	0.1273	0.0127	0.0281	0.11
Other Services	3,897	0.034	2.60	0.0884	0.2834	0.0096	0.0009	0.03
Sum	113,071	1.000		2.76		0.3148	0.0644	1.00

Table 1c: Descriptive Statistics, OES, By Industry, 1998-2007 Growth

Industry j	(1) E_j (*1000)	(2) (E_j/E)	(3) $\bar{w}_j$	(4) $(E_j/E)^* \bar{w}_j$	(5) $V_j(w)$	(6) $(E_j/E)^* V_j(w)$	(7) $(E_j/E)^* (\bar{w}_j - \bar{w})^2$	(8) Contribution to Total Variance
Mining	106	0.001	-0.02	0.0029	0.0039	0.0003	-0.0001	0.00
Retail Trade	904	-0.002	0.01	-0.0034	0.0094	0.0009	0.0033	0.11
Wholesale, Trans, Util	-774	-0.015	0.04	-0.0393	0.0118	-0.0035	-0.0008	-0.12
Construction	1,399	0.008	0.05	0.0265	0.0058	0.0023	0.0003	0.07
Manufacturing	-3,966	-0.047	0.04	-0.1288	0.0419	-0.0077	-0.0011	-0.24
Information	-20	-0.003	0.15	-0.0040	0.0412	0.0003	0.0016	0.05
Finance & Real Estate	412	-0.001	0.08	0.0029	0.0036	-0.0001	0.0008	0.02
Prof & Bus Services	4,031	0.026	0.13	0.0936	0.0303	0.0156	0.0035	0.53
Educ Services	512	0.003	0.16	0.0119	0.0944	0.0029	0.0010	0.11
Health & Social Assist	3,097	0.019	0.06	0.0608	0.0074	0.0079	0.0002	0.22
Arts & Entertainment	365	0.002	0.11	0.0066	0.0777	0.0017	-0.0003	0.04
Food & Lodging	1,841	0.009	0.03	0.0228	0.0140	0.0024	0.0044	0.19
Other Services	358	0.000	0.05	0.0017	0.0209	0.0007	0.0000	0.02
Sum	8,265	0.000		0.0541		0.0237	0.0128	1.00

Figure 5.1: Herfindahl Indices based on 3 occupational agglomerations, by establishment size category, 1998-2007

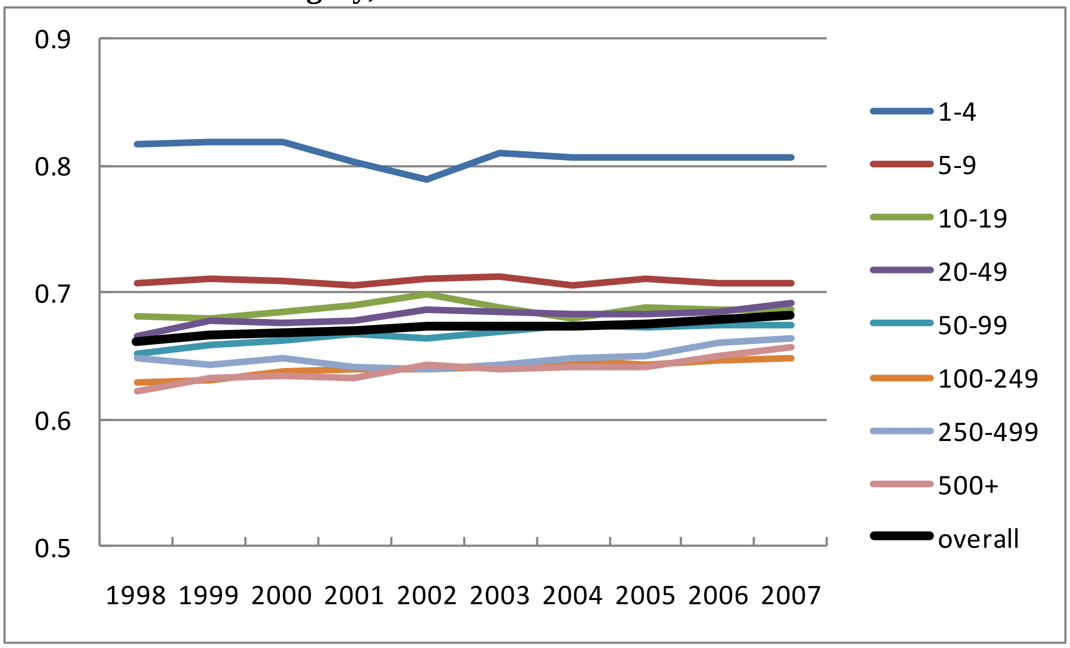


Figure 5.2: Herfindahl Indices based on 19 consistent occupational groups, by establishment size category, 1998-2007

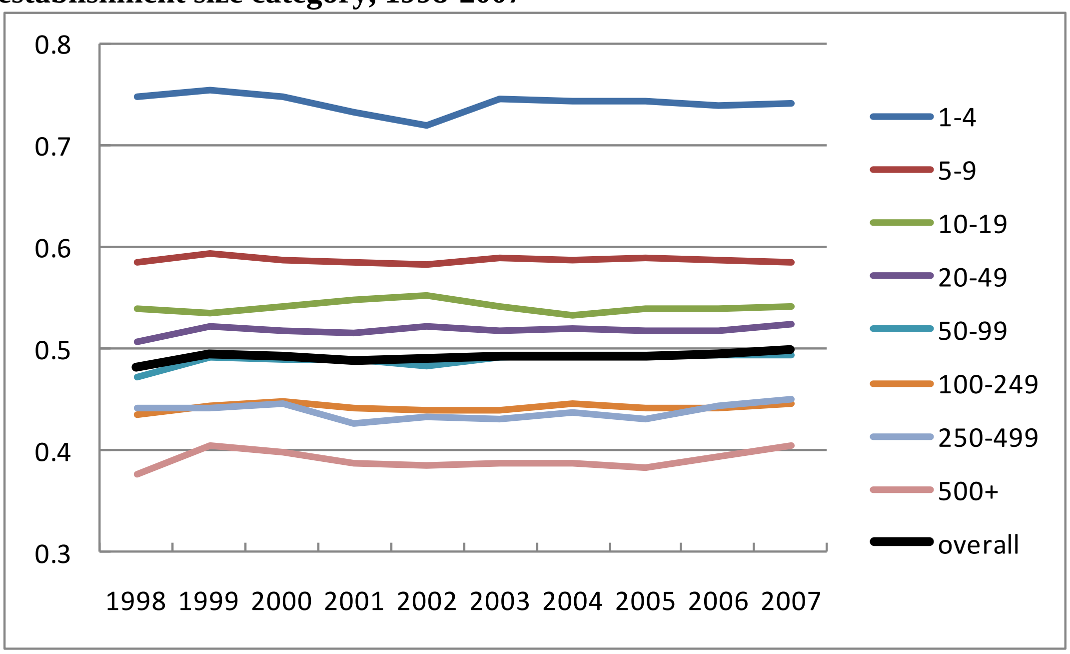


Figure 6: Variance of Log Real Hourly Wage, CPS and OES, 1996-2007

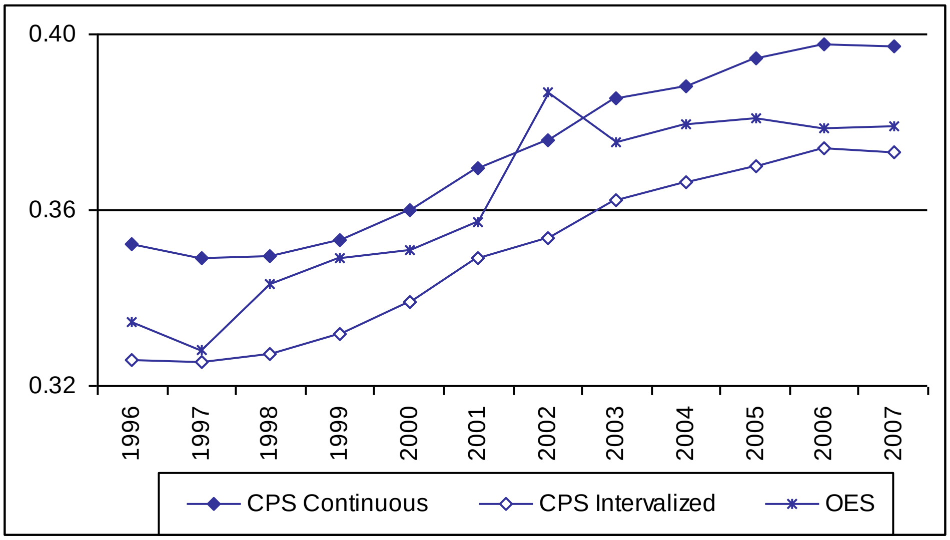
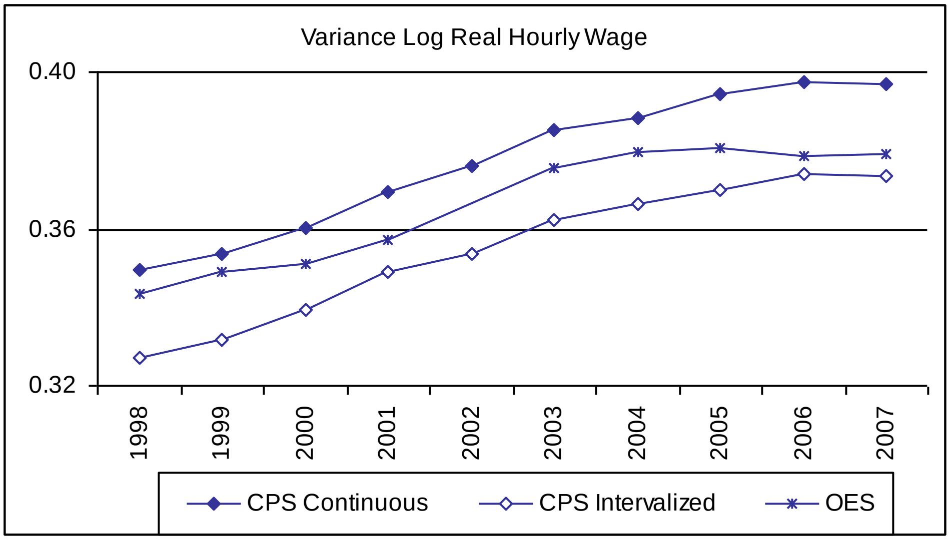


Figure 7: Variance of Log Real Hourly Wage, CPS and OES, 1998-2007
The 2002 value for the OES is linearly interpolated



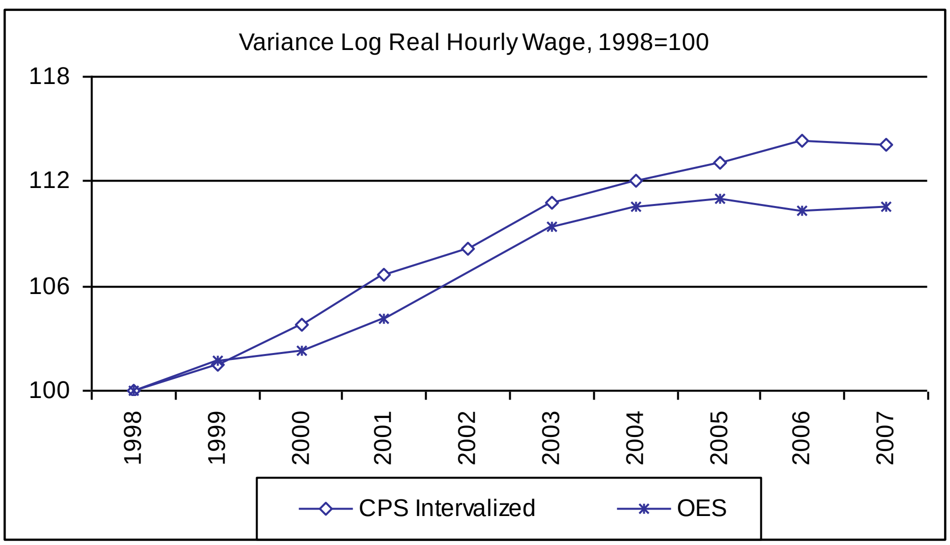


Table 2: 2007 Wage Variance and 2007 Employment by OES Wage Interval
Holding Employment Shares of Various Characteristics to their 1998 levels

		Employment (in thousands) by the 12 OES wage intervals											
	Variance	< $7.50	to $9.49	to $11.99	to $15.24	to $19.24	to $24.49	to $30.99	to $39.24	to $49.74	to $63.24	to $79.99	$80 +
1998 Comparison year	0.3432	The OES had 11 OES wage intervals in 1998, with different end points than 2007											
2007 Baseline	0.3786	9,640	16,912	16,349	17,377	14,686	12,001	9,441	6,670	4,376	2,593	1,307	1,719
Reweighting to 1998 characteristics using the method of Dinardo, Fortin, and Lemiuex (1996)													
(1) Estab size groups	0.3805	9,419	16,674	16,261	17,305	14,684	12,095	9,593	6,818	4,497	2,660	1,341	1,731
(2) Herfindahl-19occ	0.3799	9,286	16,477	16,188	17,405	14,831	12,177	9,622	6,826	4,501	2,672	1,346	1,750
(3) States	0.3769	9,816	16,900	16,368	17,418	14,686	11,984	9,428	6,622	4,327	2,550	1,279	1,693
(4) Herfindahl-3grps	0.3761	9,247	16,455	16,267	17,549	14,937	12,217	9,606	6,764	4,405	2,587	1,303	1,734
(5) OccShare groups	0.3682	9,300	16,606	16,442	17,756	15,046	12,244	9,545	6,598	4,211	2,453	1,230	1,641
(6) Industry groups	0.3738	9,180	16,503	16,337	17,640	14,982	12,215	9,555	6,658	4,386	2,629	1,319	1,668
(7) Occupation groups	0.3659	9,393	17,156	16,818	18,023	15,065	11,963	9,131	6,187	4,021	2,423	1,243	1,648
(8) = all above	0.3646	9,274	16,593	16,616	18,062	15,214	12,198	9,369	6,328	4,084	2,425	1,238	1,632
(9) = all except (7)	0.3658	9,252	16,371	16,438	17,872	15,135	12,306	9,572	6,552	4,224	2,467	1,231	1,594
(10) = (2)+(4)+(5)	0.3654	9,087	16,400	16,460	17,893	15,206	12,352	9,609	6,615	4,192	2,419	1,213	1,637
(11) = (2)+(3)+(4)+(5)	0.3646	9,240	16,395	16,465	17,909	15,189	12,329	9,593	6,584	4,165	2,394	1,197	1,621
(12) = (2)+(3)+(4)+(5)+(7)	0.3623	9,397	16,913	16,764	18,122	15,206	12,087	9,230	6,231	3,992	2,351	1,195	1,619
	Difference Explained	Employment Difference (in thousands) due to reweighting to 1998 characteristics											
(1) Estab size groups	-5%	-222	-238	-88	-72	-2	93	152	148	121	67	34	12
(2) Herfindahl-19occ	-4%	-355	-435	-161	28	145	176	182	156	125	79	40	31
(3) States	5%	176	-12	19	41	0	-17	-13	-49	-49	-43	-27	-27
(4) Herfindahl-3grps	7%	-394	-456	-82	171	251	216	165	93	29	-6	-3	15
(5) OccShare groups	29%	-340	-306	94	378	361	242	104	-72	-165	-141	-77	-78
(6) Industry groups	14%	-461	-408	-12	262	296	214	114	-12	11	36	12	-51
(7) Occupation groups	36%	-247	244	469	645	380	-38	-309	-483	-355	-171	-63	-72
(8) = all above	39%	-366	-318	268	685	529	197	-72	-342	-292	-169	-69	-88
(9) = all except (7)	36%	-388	-541	89	494	449	305	131	-118	-152	-126	-76	-125
(10) = (2)+(4)+(5)	37%	-553	-512	111	516	520	351	168	-55	-184	-175	-94	-82
(11) = (2)+(3)+(4)+(5)	39%	-401	-516	116	532	503	327	152	-86	-211	-199	-110	-98
(12) = (2)+(3)+(4)+(5)+(7)	46%	-244	1	415	745	521	86	-211	-439	-384	-243	-111	-100

Table 3: Variance of Log Real Hourly Wage, by Industry, OES, 1998-2007 Average

Industry j	Total Variance	Within Estab Variance	Between Estab Variance	Within Estab Variance	Between Estab Variance
Mining	.2818	.1575	.1244	55.9%	44.1%
Retail Trade	.2382	.1507	.0874	63.3%	36.7%
Wholesale, Trans, Util	.3161	.1784	.1377	56.4%	43.6%
Construction	.2436	.1354	.1081	55.6%	44.4%
Manufacturing	.2923	.1594	.1329	54.5%	45.5%
Information	.3543	.1872	.1671	52.8%	47.2%
Finance & Real Estate	.3925	.2217	.1708	56.5%	43.5%
Prof & Bus Services	.4430	.1773	.2657	40.0%	60.0%
Educ Services	.3269	.2298	.0970	70.3%	29.7%
Health & Social Assist	.3587	.2185	.1403	60.9%	39.1%
Arts & Entertainment	.2805	.1771	.1034	63.1%	36.9%
Food & Lodging	.1314	.0906	.0408	68.9%	31.1%
Other Services	.2811	.1327	.1484	47.2%	52.8%
All Industries	.3681	.1683	.1998	45.7%	54.3%

Figure 8: Variance of Log Real Hourly Wage, by Industry, OES, 1998-2007 Average

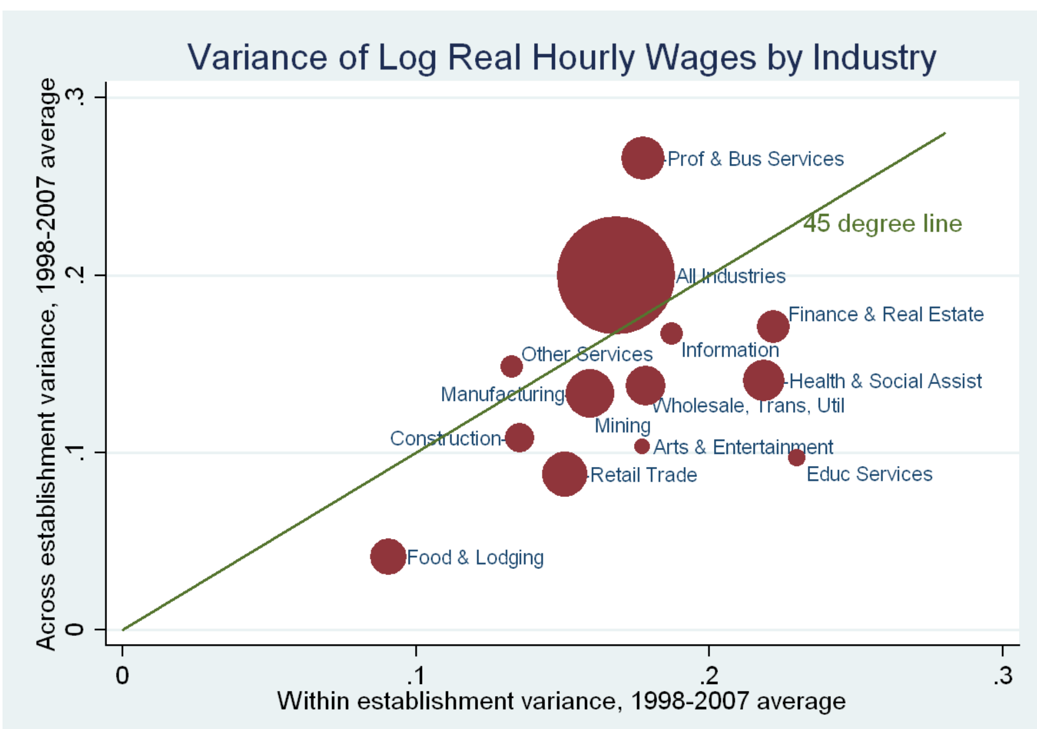


Figure 9: Variance of Log Real Hourly Wage, OES, 1998-2007

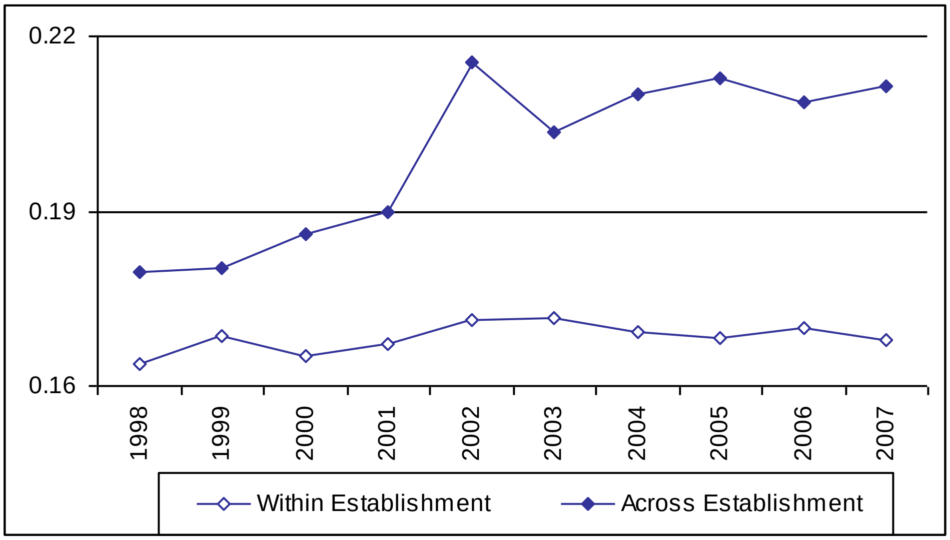


Table 4: Variance of Log Real Hourly Wage, by Industry, OES, 1998-2007 Growth

Industry j	Total Variance	Within Estab Variance	Between Estab Variance
Mining	.0039	.0030	.0009
Retail Trade	.0094	.0081	.0013
Wholesale, Trans, Util	.0118	.0020	.0097
Construction	.0058	.0023	.0035
Manufacturing	.0419	.0129	.0290
Information	.0412	.0025	.0386
Finance & Real Estate	.0036	-.0146	.0183
Prof & Bus Services	.0303	.0032	.0271
Educ Services	.0944	.0480	.0465
Health & Social Assist	.0074	-.0076	.0151
Arts & Entertainment	.0777	.0241	.0535
Food & Lodging	.0140	.0051	.0090
Other Services	.0209	.0029	.0180
All Industries (Total)	.0361	.0042	.0319

Figure 10: Variance of Log Real Hourly Wage, by Industry, OES, 1998-2007 Growth

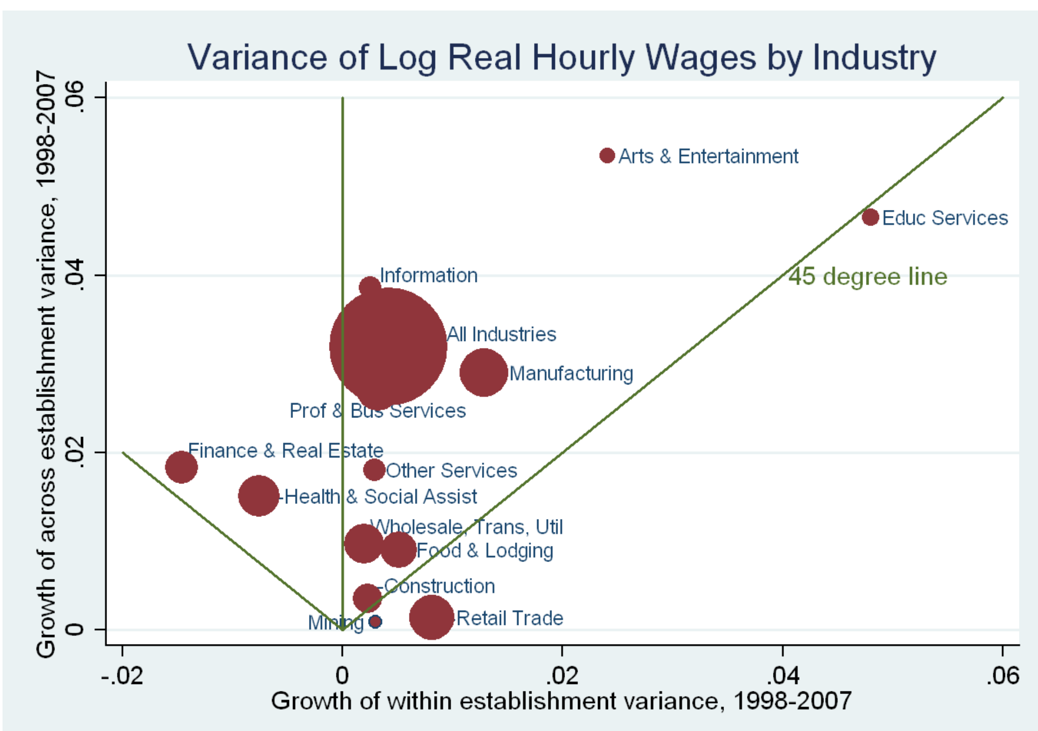


Table 5: Contribution to the Between-Establishment Variance, by Percentiles of the (weighted) Mean Establishment Log Real Hourly Wage, OES, 1998 and 2007

	1998	2007	1998-2007 Growth	1998-2007 Growth
0 - 5th percentile	.0230	.0287	.0057	17.9%
5th - 10th percentile	.0171	.0192	.0021	6.6%
10th - 50th percentile	.0409	.0462	.0053	16.6%
50th - 90th percentile	.0337	.0385	.0048	15.0%
90th - 95th percentile	.0209	.0252	.0043	13.5%
95th – 100th percentile	.0440	.0541	.0101	31.7%
Total	.1795	.2114	.0319	100.0%

Table 6: Descriptive Statistics, Full Sample and Sample of Establishments with the Highest Five Percent of Mean Establishment Wages, OES, 2007

Establishment Characteristics	Full Sample	Highest 5% Wages
Industry		
Mining	0.7%	0.9%
Construction	10.7%	0.8%
Manufacturing	12.3%	15.7%
Wholesale trade …	8.7%	4.4%
Retail trade	10.7%	4.4%
Information	3.8%	7.9%
Finance & Real Est	9.6%	12.6%
Prof & Bus Serv	18.1%	40.7%
Educ Serv	1.8%	1.7%
Health & Soc Asst	10.3%	7.3%
Arts & Entertain	2.3%	1.9%
Food & Lodging	4.1%	0.1%
Other Services	6.9%	1.8%
Establishment Size		
1-4	22.1%	36.2%
5-9	18.7%	12.8%
10-19	18.5%	13.3%
20-49	19.1%	14.8%
50-99	9.8%	8.8%
100-249	7.5%	7.6%
250-499	2.7%	3.5%
500+	1.5%	2.9%