American Sociological Review
75(3) 402–431

DOI: 10.1177/0003122410363564
http://asr.sagepub.com

Occupations and the Structure of Wage Inequality in the United States, 1980s to 2000s

Ted Mouw[a] and Arne L. Kalleberg[a]

Abstract

Occupations are central to the stratification systems of industrial countries, but they have played little role in empirical attempts to explain the well-documented increase in wage inequality that occurred in the United States in the 1980s and 1990s. We address this deficiency by assessing occupation-level effects on wage inequality using data from the Current Population Survey for 1983 through 2008. We model the mean and variance of wages for each occupation, controlling for education and demographic factors at the individual level to test three competing explanations for the increase in wage inequality: (1) the growth of between-occupation polarization, (2) changes in education and labor force composition, and (3) residual inequality unaccounted for by occupations and demographic characteristics. After correcting for a problem with imputed data that biased Kim and Sakamoto's (2008) results, we find that between-occupation changes explain 66 percent of the increase in wage inequality from 1992 to 2008, although 23 percent of this is due to the switch to the 2000 occupation codes in 2003. Sensitivity analysis reveals that 18 percent of the increase in inequality from 1983 to 2002 is due to changes in just three occupations: managers "not elsewhere classified," secretaries, and computer systems analysts.

Keywords

labor markets, trends in inequality, occupational polarization, human capital models

Sociologists have long regarded occupations as the key indicators of workers' positions in the societal division of labor. The widespread use of measures of occupational prestige, status, and average earnings to represent the structure of inequality assumes that occupations are the basic components of stratification systems in industrial nations (Parkin 1971). Many occupations constitute distinct labor markets within which supply and demand for particular types of labor intersect, and so occupational differences help explain how inequality is generated (Bielby and Kalleberg 1981; Spilerman 1977; Stolzenberg 1975).

The centrality of occupations to sociologists' theoretical conceptions of the stratification system suggests that changes in the occupational structure—that is, in the relative size and wages of different occupations—would play a prominent role in sociological

[a]University of North Carolina-Chapel Hill

Corresponding Author:
Ted Mouw, University of North Carolina, Department of Sociology, CB#3210, 155 Hamilton Hall, Chapel Hill, NC 27599 (http://www.tedmouw.org)
E-mail: tedmouw@email.unc.edu

attempts to explain the well-documented increase in wage inequality in the United States over the past 25 years. Indeed, popular accounts of the increase in inequality often assume a growing polarization of occupations, illustrated by changes such as the loss of high-wage, blue-collar occupations through deindustrialization; the increase in poorly paid service-sector jobs; and the rapid growth in wages of managerial and professional occupations. Surprisingly, systematic research on the role of occupations in generating patterns of wage inequality is scarce (DiPrete 2007; Morris and Western 1999; Myles 2003). The few studies that explicitly examine how occupations generate economic inequality appear to confirm the assumption that differences between occupations are primarily responsible for the growth of wage inequality (e.g., Autor, Katz, and Kearney 2006; Massey and Hirst 1998; Weeden et al. 2007; Wright and Dwyer 2003). Recently, however, several sociologists have challenged this view of the centrality of the occupational structure in explaining changes in inequality. Kim and Sakamoto (2008) use data from the Current Population Survey (CPS) to examine trends in between- and within-occupation inequality from 1983 to 2002. They find that the within-occupational component of inequality increased sharply during their study period, while the between-occupation component remained relatively constant (see also Raffalovich 1993). Kim and Sakamoto (2008: 153) conclude "that the traditional sociological focus on occupations may be inadequate as the primary explanation of rising wage inequality" and suggest that occupations are declining in significance and becoming less influential in the New Economy. This conclusion has serious implications for sociologists' assumptions about the central role played by occupations in the process of stratification.

In this article, we reevaluate the role of occupations in explaining the increase in wage inequality. We use data from the Current Population Surveys from 1983 to 2008 to estimate models that simultaneously test individual- and occupation-level effects on changes in the mean and variance of wages. This allows us to analyze the impact of occupational structure on trends in inequality while controlling for changes in the returns to education and the demographic composition of the labor market. In addition, we present results at the aggregate- and detailed-level of occupations. In so doing, we estimate the contribution of specific occupations to the increase in inequality during this period, adjusting for changes in individual-level variables. Finally, we correct for the Census Bureau's treatment of missing wage data, which imputes missing wages based on non-missing data from aggregate occupational categories rather than detailed occupations, thereby mechanically increasing within-occupation inequality. We show that much of the trend in within-occupation inequality reported by Kim and Sakamoto (2008) is an artifact of this incorrect approach to handling missing data.

EXPLAINING THE INCREASE IN WAGE INEQUALITY IN THE UNITED STATES

Debates on trends in wage inequality typically identify three main sources. First, sociologists (but usually not economists) generally assume that wage differences between occupations are important for generating wage inequality. Second, economists—and many sociologists—argue that the growth of wage inequality reflects primarily changes in the composition of the labor force with regard to human capital characteristics such as education, experience, and other skills. Finally, a growing literature in economics (as well as in sociology [e.g., Kim and Sakamoto 2008]) maintains there is residual inequality in wages after accounting for occupations and labor force characteristics.

Between-Occupation Inequality

Sociologists often argue that occupational structure (i.e., the relative size and average wages of different occupations) is central to explanations of inequality (e.g., Bielby and Kalleberg 1981; Blau and Duncan 1967; Featherman and Hauser 1978; Treiman 1977). Recent descriptions such as disaggregate structuration (e.g., Grusky and Sørensen 1998; see also Weeden et al. 2007) contend that members of detailed occupational categories are fairly homogenous in terms of their life chances. Economic inequalities should thus lie primarily between occupations, not within them. The validity of the assumption that occupations are integral to explanations of economic inequality depends on the degree to which wages differ between, relative to within, occupations.

Theoretically, between-occupation wage differences, and changes in average occupational wages, result from several sources. First, occupations vary in their skill, that is, the degree of complexity of occupational activities and the amount of training time required to perform them adequately. Other factors, such as average education level and occupation-specific skills not captured by broad educational categories, may also reflect an occupation's skill level. Between-occupation wage differences may also be represented by differences in the returns to education and other labor force characteristics, such as gender, that vary by occupation. In our empirical analysis, we test for occupation effects beyond the returns to labor force composition by estimating models that include individual-level controls for education and other wage-related individual characteristics.

Second, occupations differ in the extent to which they adopt and maintain institutional mechanisms of social closure such as licensing, educational credentialing, voluntary certification, association representation, and unionization (Weeden 2002). These institutional mechanisms of social closure increase earnings by restricting the supply of workers, increasing demand and channeling it to the occupation, and signaling the quality of service. Weeden's (2002) analysis found that occupational differences in technical complexity and measures of social closure affect occupational earnings.

Recent research has found evidence of between-occupation effects on wage inequality, suggesting that the occupational structure has become increasingly polarized: high- and low-paying occupations have grown in size, especially during the 1990s. Massey and Hirst (1998) found empirical support for an hourglass pattern of growth in high- and low-paying occupations from 1969 to 1989 for men (but not for women). This is a change from the escalator pattern of 1949 to 1969, in which people worked in progressively higher paying occupations. Wright and Dwyer (2003) examined the growth and decline of jobs of varying quality during U.S. economic expansions since the 1960s. They conclude that jobs grew in the top and bottom quintiles in the 1990s but shrank in the middle quintile.[1] Autor and colleagues (2006) show that the share of total hours worked by members of high- and low-wage occupations increased during the 1990s, while the employment shares of middle-wage occupations declined (for a similar analysis of the United Kingdom, see Goos and Manning 2007). Finally, Weeden and colleagues (2007) found that much of the growth in inequality during this period was due to differences across "big classes" (aggregate groups of occupations), especially for men. At the same time, they recognize that the absence of controls for education in their analysis may confound their interpretation of between-occupation changes in inequality with changes in the returns to education, an issue to which we now turn.

Human Capital and Labor Force Composition

Might the growth in wage inequality be due to changes in the returns to or the levels of education and skill in the labor force, not from changes in occupational structure? Indeed, increases in the returns to education, as exemplified by the widening college/high-school wage gap (Autor, Katz, and Kearney 2008; Katz and Autor 1999), could explain much of the polarization of the labor force into good and bad jobs. Studies show considerable support for this notion. During the 1990s, the theory of "skill-biased technical change" (SBTC) was the most prominent explanation of rising inequality in the economics literature. According to this theory, fundamental technological advances that are "biased" in favor of high-skilled workers, such as the use of computers, drive recent changes in inequality (Acemoglu 2002; Katz and Murphy 1992; Levy and Murnane 1992; for a recent review, see Lemieux 2008). An increasing demand for skilled workers driven by technological changes would help explain the deterioration of wages for less-skilled workers during the 1980s and the growth in demand for workers with high education (Levy and Murnane 1992).

Or, is the increase in inequality due to changes in labor force composition, not technological changes per se? It is well known—and well grounded in human capital theory (see Mincer 1974)—that more educated and more experienced workers tend to have higher variances in wages. As a result, an increase in the average educational level and experience (i.e., age) of the labor force tends to increase wage inequality (Lemieux 2006a). With respect to trends in occupation-level inequality, this composition perspective implies the same empirical prediction as SBTC: observed measures of human capital, such as education and experience, should attenuate or explain away trends in between-occupational inequality.

Given the possibility that returns to skill or changes in labor force composition could account for the impact of changes in occupational structure on inequality, we incorporate the human capital model into our empirical estimates of occupational inequality. We present individual-level models of the mean and variance of different occupations, controlling for education and potential labor market experience as well as other individual characteristics that may explain the growth of inequality, such as gender, race, and union membership.

Residual Inequality

A third explanation argues that the rise in wage inequality cannot be explained by either changes in occupational structure or observed human capital variables. The recent economics literature on rising inequality pays considerable attention to the decomposition of trends in inequality into the effect of observed labor force characteristics and an unexplained, "residual" component not explained by human capital variables (Acemoglu 2002; Autor et al. 2006; Card and DiNardo 2002; Juhn, Murphy, and Pierce 1993; Katz and Autor 1999; Lemieux 2006b). The concept of residual inequality in human capital models has a direct parallel with the measure of within-occupation inequality that Kim and Sakamoto (2008) use in examining occupational inequality (i.e., the component of inequality not explained by occupational structure, or at least by a particular set of occupational codes).

Examining trends in residual inequality helps adjudicate between SBTC and alternative theories (Acemoglu 2002; Card and DiNardo 2002; Juhn et al. 1993; Lemieux 2006a). SBTC predicts that inequality will increase within and between education levels. Within-education inequality increases because of the greater demand for skills, some of which are unobserved to the researcher and vary among workers with the same level of education. Workers who best adapt to the new technologies earn a wage premium, leaving behind less adaptable workers with the same education (Acemoglu 2002; Juhn et al. 1993; Lemieux 2006a). Overall, the rise of residual inequality in the 1970s and 1980s is consistent with SBTC's prediction: Katz and Autor (1999) find that residual inequality accounts for around 60 percent of the increase in inequality between 1963 and 1995; Juhn and colleagues (1993) and Acemoglu (2002) find significant increases in residual inequality since the late 1970s after controlling for individual-level predictors of wages such as education, experience, and demographic categories. In a prominent sociological account of increasing inequality, Bernhardt and colleagues (2001) note that more than half of the increase in inequality for men has been within groups of workers with the same age, education, and experience.

A recent revisionist argument challenges the SBTC explanation by highlighting two shortcomings based on trends in residual inequality during the 1990s. First, residual inequality grew sharply during the 1980s but then slowed, or ceased, during the 1990s (Card and DiNardo 2002). Second, Lemieux (2006a) shows that after adjusting for changes in the proportion of workers in different education and experience categories, the trend in residual inequality after 1990 disappears. This suggests that recent increases in inequality are due to changes in the demographic composition of the labor force rather than an underlying process of skill-biased technological change.

The debate between the SBTC and revisionist interpretations is relevant for our discussion of occupational inequality for two reasons. First, as noted earlier, there is a conceptual parallel between residual inequality in human capital models and within-occupation inequality. Consequently, studies indicating little or no increase in residual inequality during the 1990s (Card and DiNardo 2002; Lemieux 2006a) raise questions about the trend in residual within-occupational inequality found by Kim and Sakamoto (2008) during the same period.

Second, the economics literature is missing a careful consideration of the role of occupational structure (Autor and colleagues [2006] and Goos and Manning [2007] are exceptions). Given the focus on observed and residual components of inequality, this is not a trivial omission. For example, one reason why the variance of wages increases with education (Lemieux 2006a) is that specialized training encourages workers to move into heterogeneous occupations. As a result, changes in occupational structure may explain an increase in within-education residual inequality or average wages. Weeden and colleagues (2007) suggest that a growing demand for high-skill occupations may explain some of the increase in residual inequality in human capital models. On the other hand, the reverse could also be true: an increase in the returns to higher education might explain trends in between-occupation inequality. In this case, it is not the occupational structure per se that generates inequality, but a change in the fundamental relationship between education and earnings.

Three Competing Explanations

Based on the preceding discussion, we test the following competing explanations of the recent rise in wage inequality:

(1) *Changes in occupational structure:* This is the between-occupation argument predicting that changes in the average wage and size of occupations explain the increase in inequality.
(2) *Education and labor force changes:* Changes in the returns to human capital and the demographic composition of the labor force explain the increase in inequality.
(3) *Residual inequality, net of controls for occupation and observed human capital variables:* This combines the explanations based on trends in residual inequality in human capital models with Kim and Sakamoto's (2008) argument about trends in within-occupation inequality.

Our models for the mean and variance of occupational wages control for education and basic demographics at the individual level, allowing us to test these competing hypotheses simultaneously.

METHODS

Variance Decomposition

We now present descriptive methods to calculate trends in between- and within-occupation inequality as well as multivariate fixed-effects models that we use to test the competing hypotheses. Our measure of inequality is the variance of log wages (varlog), which is used in many recent studies of trends in U.S. inequality (e.g., Lemieux 2006). Varlog is particularly attractive for our purposes because it is easily decomposed by groups into within- and between-occupation inequality. The variance of log wages can be decomposed by occupation using the standard decomposition of variance:

$$\begin{aligned} Var[\ln wage] &= Var[\ln wage|occ] \\ &= Var[E[\ln wage|occ]] \\ &+ E[Var[\ln wage|occ]] \end{aligned} \tag{1}$$

In the right-hand side of Equation 1, the first term (the variance of mean occupation wages) is the between-occupation component. The second term is the within-occupation component. For each occupation, three parameters are needed to calculate the overall level of inequality: the proportion of workers who work in occupation j at time t, p_{jt}, the mean log occupational wage, μ_{jt}, and the within-occupation variance of log wages, σ^2_{jt}. Using p_{jt}, μ_{jt}, and σ^2_{jt}, Equation 1 can be rewritten as follows:

$$\begin{aligned} Var[\ln wage|occ] &= \sum_j p_{jt}(\mu_{jt} - \bar{\mu}_t)^2 \\ &+ \sum_j p_{jt}\sigma^2_{jt} \end{aligned} \tag{2}$$

where $\bar{\mu}_t$ is the overall mean log wage at time t.

In Equation 2, note that either changes in occupational size, p_{jt}, or average occupational wages, μ_{jt}, can affect between-occupation inequality. To differentiate the effect of occupational size and wages, Equation 2b presents a decomposition of between-occupation inequality:

$$\begin{aligned} \sum_j p_{jt}(\mu_{jt} - \bar{\mu}_t)^2 &= \sum_j p_j^{1983}(\mu_{jt} - \bar{\mu}_t)^2 \\ &+ \sum_j (p_{jt} - p_j^{1983})(\mu_{jt} - \bar{\mu}_t)^2 \end{aligned} \tag{2b}$$

where p_j^{1983} is the size of occupation j in 1983, the first year of our data. The first term on the right-hand side of Equation 2b depicts between-occupation inequality based on current average wages, holding the occupational distribution to its 1983 levels. The second term is the component of between-occupation inequality due to changes in the size of occupations.[2]

Modeling Occupational Inequality

We use Equations 1 and 2 to divide overall inequality into between- and within-occupation components. The equations do not, however, control for individual-level factors that affect inequality and relate to the sorting of workers into different occupations. In particular, we are concerned about the effects of education and changes in the composition of the labor force. As Hypothesis 2 suggests, increasing returns to education, such as the rise of the college premium (Katz and Autor 1999), increase the wages of occupations that require college degrees or higher, thereby increasing between-occupation inequality.[3]

To simultaneously estimate the effect of occupations and the controls on inequality, we model the mean and variance of log wages using an OLS model with nonconstant variance (for a recent discussion of variance models in inequality research, see Western and Bloome 2009). We do this in three steps.

(1) We begin with a model for log wages that includes occupation-fixed effects:

$$\ln wage_{ij} = \beta X_{ij} + \alpha_j occ_i^j + \varepsilon_{ij},\ \varepsilon_{ij} \sim N(0,\sigma_{ij}^2) \tag{3a}$$

where ln $wage_{ij}$ is the log wage of individual i in occupation j, X is a set of individual-level controls (dummy variables for 5-year intervals of age, each category of education, race, and union membership), occ_i^j is an occupation dummy variable (one for each detailed 3-digit occupation j in the data), α_j is the fixed effect of occupation j, and ε_i is a heteroskedastic error term. Because of the large number of occupation fixed effects, we first transform the data to remove the fixed effects and estimate a model of deviations from the means of each variable:

$$\begin{aligned}\ln wage_{ij} - \ln wage_{.j} = \beta(X_{ij} - \overline{X}_{.j}) \\ + (\varepsilon_{it} - \bar{\varepsilon}_{.j})\end{aligned} \tag{3b}$$

where ln $wage_{.j}$ is the average wage in occupation j, and $\bar{X}_j$ represents the average of each control in occupation j. After estimating 3b, we recover the occupation fixed effects with Equation 3c (see Bollen and Brand forthcoming; Wooldridge 2002):

$$\alpha_j = \ln wage_{.j} - \hat{\beta}(X_{ij} - \bar{X}_{.j}) \tag{3c}$$

Because point estimates of OLS coefficients are consistent in the case of heteroskedastic errors, we obtain unbiased estimates of the parameters in Equations 3b and 3c.[4] We use the estimated fixed effects (α_j's) as the mean occupational wage net of controls. Because the within-occupation component of inequality is the variance of mean occupation wages (see Equation 2), the (weighted) variance of the fixed effects tells us how much between-occupation inequality is generated in the data holding the controls constant.

(2) In Step 2, we use the coefficient estimates and occupation mean fixed effects from the first step to calculate the square of the residual for each case; we use this as the dependent variable for our model of the variance:

$$\begin{aligned}e_{it}^2 &\equiv [\ln wage_{ij} - (\hat{\beta}X_{ij} + \alpha_j occ_i^j)]^2 \\ &= \phi_t X_{it} + \delta_j occ_i^j + \omega_{ij}\end{aligned} \tag{3d}$$

We estimate this model and recover the occupation fixed effects using the same approach as in Equations 3b and 3c. The occupation variance fixed effects in Equation 3d, δ_j, represent the estimated variance of log wages in each occupation, controlling for covariates such as education and age that also affect the variance.

(3) Finally, we use the estimated mean and variance fixed effects for occupations generated in Steps 1 and 2 to calculate the estimates of between- and within-occupation inequality using Equation 4:

$$\begin{aligned}Var[\ln wage|occ,X] = \sum_i{}' p_j(\alpha_j - \bar{\alpha})^2 \\ + \sum_i p_j \delta_j\end{aligned} \tag{4}$$

We estimate Equations 3a and 3d separately by gender because of the degree of occupational segregation between men and women, and the possibility of different returns to education by gender. This allows the effect of gender on the mean and variance of wages to vary by occupation. Before estimating the overall occupational inequality in Equation 4, we combine the gender occupational fixed effects for each occupation. The combined mean wage fixed effect of the occupation is the average of the male and female fixed effects, weighted by the number of workers of each gender. The combined within-occupation variance fixed effect for occupation j is (using Equation 1) the mean of the (male and female) variance fixed effects, δ_j, plus the variance of the (male and female) mean fixed effects, α_j, both weighted by the relative proportion of men and women in occupation j. The final term (the variance of the mean fixed effects) is the degree of within-occupation inequality generated by the gender wage gap within occupations, which we report separately in the results below.

We have controlled for the effect of education and demographic variables on the mean and variance of occupational wages, but we have not adjusted the size of occupations for changes in these variables. We assume that the demand for specific types of labor drives changes in the relative size of occupations and thus represents the occupational structure of a given society. At a deeper level, however, it is a classic chicken-or-egg question: Did the structure of occupations change because of a change in the supply of workers with different education credentials? Or, did average education levels change due to a changing demand for specific types of workers? One approach to this problem would use changes at the industry level (i.e., at the level of production of different goods and services) as a measure of demand fluctuations and use the average share of occupations by industry to estimate the change in occupational composition due to demand side effects. We leave this more complex approach to future research. We note, though, that if the change in inequality resulted entirely from education, there would be no occupational-level fixed effects after controlling for the wage effect of education. As a result, in this study we are estimating the effect of the relative size of different occupations combined with differences in wages not explained by other explanatory variables.

Calculating the Effect of Specific Occupations on Inequality

In the final part of the article, we attempt to estimate the effects of specific occupations on the overall level of wage inequality. To do this, we first estimate the counterfactual by substituting the time $= t - 1$ values of p (the size of the occupation), μ (average wage), and σ^2 (the within-occupation variance in wages) for occupation i while using the time $= t$ values for the rest of the occupations.[5] In other words, the counterfactual is: What would the overall level of inequality be if everything changed except for the size and wages of occupation i? The change in wage inequality attributable to occupation i is the difference between the actual level of inequality and the counterfactual:

Change in inequality due to occupation i =

$$\text{var}(\ln wage_{time\,=\,t}) - \text{var}(\ln wage_{time\,=\,t} | p_{it} = p_{i,t-1},\ \mu_{it} = \mu_{i,t-1},\ \sigma^2_{it} = \sigma^2_{i,t-1}) \qquad (5)$$

We estimate this twice for each occupation. First, we use the actual values of p, μ, and σ^2; second, we substitute the mean and variance fixed effects estimated in Equations 3a and 3d (i.e., the adjusted mean wage and variance of each occupation, net of individual-level variables).

Table 1 presents a simple example of this calculation. In this example, overall wage inequality is 2 at time 1, and 7 at time 2. The increase in inequality is due to a decrease in the mean log wage of occupation A and an increase in the within-occupation inequality of occupation B. Table 1 shows that the time $t = 2$ inequality would have been 5 if occupation B were held to its time $t = 1$ levels; hence, the increase in inequality attributable to occupation B is $7 - 5 = 2$.

In general, the effect of changes in p_{it}, μ_{it}, or σ^2_{it} on overall inequality is easy to depict. An

Table 1. Example of Inequality Calculations

	Time 1			Time 2		
Occupation	$p_{i,1}$	$\mu_{i,1}$	$\sigma^2_{i,1}$	$p_{i,2}$	$\mu_{i,2}$	$\sigma^2_{i,2}$
A	.5	2	1	.5	0	1
B	.5	4	1	.5	4	5

Note: Time $t = 1$ inequality:
$Var(\ln wage_{time\,=\,1}) = Var(\mu_i) + mean(\sigma_i^2) = [.5(2-3)^2 + .5(4-3)^2] + [.5*1 + .5*1] = 1 + 1 = 2$
Time $t = 2$ inequality:
$Var(\ln wage_{time\,=\,2}) = Var(\mu_i) + mean(\sigma_i^2) = [.5(0-2)^2 + .5(4-2)^2] + [.5*1 + .5*5] = 4 + 3 = 7$
Counterfactual: $\text{var}(\ln wage_{time\,=\,2} | occ\_B\_unchanged) = [.5(0-2)^2 + .5(4-2)^2] + [.5*1 + .5*1] = 4 + 1 = 5$ $\Delta inequality | due\_to\_B = 7 - 5 = 2$

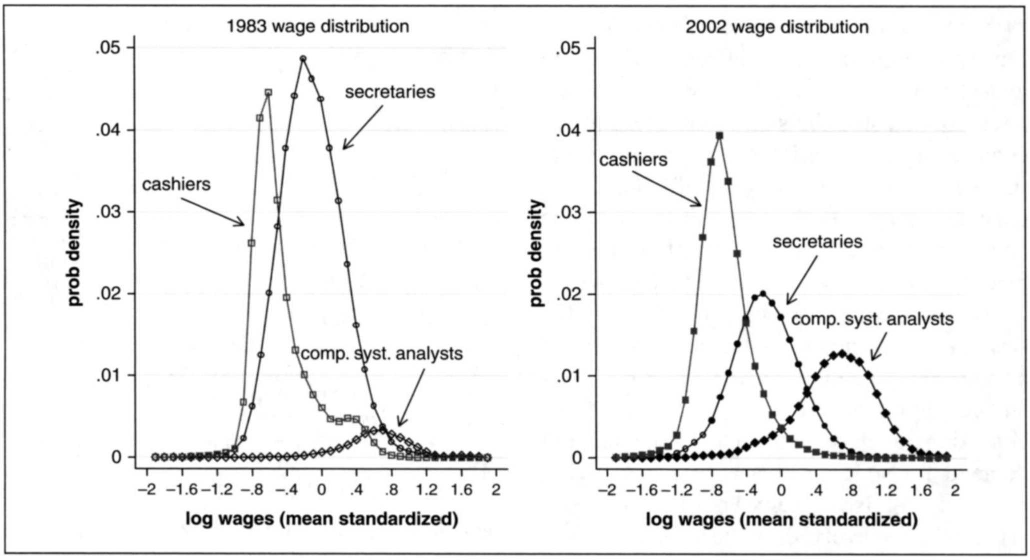

Figure 1. Changes in the Wage Distribution by Occupation, 1983 to 2002
Note: Kernel density plots. Height indicates occupation size.

increase in size (p_{it}) of a middle-wage occupation will decrease inequality, with the opposite being true for increases in low- or high-wage occupations. Increasing average wages for a high-wage occupation, or decreasing wages for a low-wage occupation, will increase inequality. Increasing within-occupation inequality always increases overall inequality, regardless of an occupation's position in the occupation wage distribution. Equation 5 is simply a way to quantify the magnitude of these changes on the overall level of inequality.

Figure 1 illustrates changes in these three components for three detailed occupations: secretaries, cashiers, and computer systems analysts. We plot the probability density of the distribution of log wages for each occupation in 1983 to 1985 and 2000 to 2002. As in the analysis below, all log wages are mean standardized by subtracting the overall mean log wage for each year.[6] The location of the middle of the distribution on the X-axis indicates average wages, the overall height of the distribution depicts the size of the occupation, and within-occupation inequality can be visualized by how "spread out" the distribution is. For example, secretaries are a large, middle-wage occupation in 1983, but the occupation experiences a dramatic decline in size by 2002. A decline in size of a middle-wage occupation increases overall inequality, everything else being equal, and this can be visualized using Figure 1. If one calculates the combined wage distribution of these three occupations in 1983 to 2005 by adding the probability densities, a decrease in the number of secretaries would increase the relative size of the low (cashiers) and high (computer system analysts) wage occupations, resulting in an increase in the variance. In contrast to secretaries, cashiers' wage distribution shifts to the left (indicating a decline in relative wages) and spreads out (indicating an increase in within-occupation inequality). Finally, computer systems analysts are a rapidly growing high-wage occupation, which would lead to an increase in between-occupation inequality, everything else being equal.

The decomposition of variance in Equation 5 is not additive. A change in p_i or μ_i affects the overall mean wage and hence the variance of mean occupational wages. As a result, the sum of changes in inequality attributable to each occupation separately will not necessarily add up to the actual change in inequality between two time periods. Instead, the decomposition should be thought of as the hypothetical change in inequality due to occupation *i*, holding everything else constant at the current levels.

DATA

We use data from the merged outgoing rotation groups (ORG) of the Current Population Survey (CPS) from 1983 to 2008. The online supplement (http://asr.sagepub.com/supplemental) presents more detail on our sample, including the Stata computer code necessary to replicate it.[7] "Outgoing rotation groups" refers to households that are in the CPS panel for four months, rotate off for four months, and then are back on for four more months. Wage data are collected only in the fourth and eighth months of participation, when respondents rotate out of the sample. Many researchers use the CPS-ORG to study trends in wage inequality (e.g., Card and DiNardo 2002; Kim and Sakamoto 2008; Lemieux 2006a; Weeden et al. 2007). We follow established conventions by restricting the sample to workers between the ages of 18 and 65 years who reported hourly wages between $1 and $100 in 1979 dollars ($2.69 and $269 in 2005 dollars). Wage information is collected only for workers who are not self-employed.[8] We convert all wage data to inflation adjusted 2005 dollars. As described in the online supplement, we merge six 1980 occupation categories that are combined in the 1990 Census codes, and drop one (waitresses, see note 8), leaving us with 496 3-digit occupations. All the results we present weight the data using the sampling weights multiplied by usual hours worked. Weighting by hours worked allows the inclusion of part-time workers without counting them as full-time workers (Lemieux 2006a). We exclude the years 1994 and 1995 from the analysis because of the lack of documentation indicating whether wages were imputed.[9]

To protect privacy, the CPS truncates wages at the top of the distribution by top coding. In addition, the value of the top code has changed over time.[10] Failure to adjust for top-coded wages results in a substantial increase in inequality between 1988 and 1989, when the top-code level increased. We follow standard practice (e.g., Card and DiNardo 2002; Weeden et al. 2007) and replace top-coded wages with 1.4 times the top-coded value.[11] Because wages are top coded, the CPS is not capable of analyzing inequality in the upper tail of the income distribution.[12]

Table 2 shows trends in inequality and missing data in the CPS from 1983 through 2008. Column 3 shows time trends in the proportion of cases that are top coded in the CPS data. As noted earlier, there is a big change in the proportion of cases top coded between 1988, when 5.5 percent of the cases are top coded, and 1989, when .6 percent are top coded. In recent years, however, the proportion of top-coded cases has dropped to about .1 percent of the overall sample.

The CPS used a consistent set of 1980 Census occupational codes for the 1980 to 2002 data.[13] In 2003, the CPS switched to the 2000 Census occupational codes, which do not map directly onto the 1980 codes. This would break our time series on occupational inequality into two parts, 1983 to 2002 and 2003 to 2008. Fortunately, the Census Bureau released a set of CPS data from 2000 to 2002 that includes the new and old occupational codes. We use this double-coded occupation data to estimate between- and within-occupation inequality for both sets of occupation codes for 2000 to 2002 to determine how much the switch in codes affects the decomposition of occupational inequality, allowing us to piece together the two time series (see Table 3).

Missing Data

Column 4 of Table 2 shows time trends for the proportion of cases that have wages

Table 2. Trends in Inequality and Missing Data in the CPS, 1983 to 2008

					Wages of Missing Cases Using Matched CPS Data[a]		
	1	2	3	4	5	6	7
Year	Varlog	Number of Non-missing Cases	Proportion Top Coded	Proportion Imputed	Percent Non-missing Matched[b]	Percent Missing Matched	Difference in Variance[c]
1983	.2865	138,544	.023	.143	.298	.176	.0193
1984	.2957	140,162	.028	.152	.442	.239	.0134
1985	.3032	143,093	.034	.145	.220	.124	.0324
1986	.3116	148,304	.040	.107	.366	.214	.0125
1987	.3152	144,614	.047	.137	.571	.337	.0182
1988	.3189	136,193	.055	.156	.565	.340	.0202
1989	.3095	136,583	.006	.169	.568	.356	.0111
1990	.3084	142,745	.008	.169	.564	.349	.0090
1991	.3059	138,602	.009	.169	.583	.359	.0121
1992	.3069	136,124	.010	.171	.580	.355	.0087
1993	.3116	132,354	.011	.184	.603	.461	.0323
1996	.3227	100,732	.018	.244	.375	.212	−.0092
1997	.3222	102,764	.021	.250	.556	.306	−.0103
1998	.3206	102,199	.008	.261	.547	.302	−.0005
1999	.3238	99,008	.010	.297	.544	.297	−.0131
2000	.3292	97,012	.011	.321	.524	.280	.0021
2001	.3329	100,928	.012	.338	.511	.276	−.0168
2002	.3371	109,493	.014	.337	.492	.252	−.0026
2003	.3252	104,846	.001	.351	.515	.261	−.0044
2004	.3276	103,366	.001	.347	.501	.251	−.0051
2005	.3315	105,065	.001	.343	.498	.246	.0033
2006	.3318	104,896	.001	.345	.524	.262	.0016
2007	.3295	105,393	.001	.337	.530	.265	.0084
2008	.3305	104,225	.001	.334	.270	.126	−.0095

[a]For details on the longitudinal matching of CPS data see the text.
[b]The proportion of individuals with wage data at time t that can be matched to their corresponding $t-1$ or $t+1$ wages.
[c]The variance of ($t-1$ or $t+1$) log wages in Column 6 minus the variance of ($t-1$ or $t+1$) log wages in Column 5.

imputed by the Census Bureau. The proportion of imputed cases rose from .143 in 1983 to .334 in 2008. The treatment of missing wage data is an important issue for the difference between our results and those reported by Kim and Sakamoto (2008), so we consider several different approaches to handling missing data.

Matched CPS data. One advantage of the CPS is that, as described earlier, respondents appear in the outgoing rotation groups twice, separated by 12 months. As a result, respondents who do not change households can be matched across successive years. Although there is no individual identification consistent across years of the CPS data, respondents can be matched on the basis of household identification, line number, sex, and race (for more details, see Madrian and Lefgren 2000; Welch 1993). After selecting individuals who are potential matches, we discard those where the age difference is greater than two years or less than one year (depending on the exact date of the interview, respondents could age zero, one, or two years over the 12-month period).

By matching respondents with valid and missing wage data in our sample to their

$t - 1$ or $t + 1$ CPS wage record, we can make some inferences about the wages of cases with missing data in our sample. Columns 5 and 6 of Table 2 show the proportion of cases with valid and missing wage data at time t that we successfully matched to wage data at time $t - 1$ or time $t + 1$. Although the overall match rate for cases using this procedure is about 70 to 75 percent (results available on request), in Column 5 the match rate for wages is only 50 to 60 percent (in some valid matches the respondent is not in the labor force or has missing data). Some years have lower match rates because the sample frame changed and different households were selected (for details, see Madrian and Lefgren 2000). In Column 6, we see that the match rate for cases with missing wage data at time t (i.e., in our sample) is about half the match rate for cases with valid wage data; this is not surprising as there may be systematic reasons why the data are missing.

Column 7 of Table 2 presents time trends in average difference between the variance of the matched ($t - 1$ or $t + 1$) log wages for cases with missing and valid wage data in our sample at time t. Positive values indicate higher average variance for cases with missing (time t) wage data. Notably, the difference in the variance is small (compared with the average varlog in Column 1). In addition, there is no discernable time trend: Column 7 provides no evidence that the variance of log wages for cases with missing data at time t has increased over time relative to non-missing cases. It is, of course, possible that something is going on with the cases that cannot be linked to valid CPS wage data using this matching algorithm. As a result, we next consider alternative approaches that attempt to impute the wage data for missing cases.

Hot-deck imputation. For workers with missing wages, the Census Bureau imputes wages using a hot-deck procedure that replaces missing wages with those of the last processed worker with valid wage data and the same values on a set of match variables. The variables used for matching are education, sex, race, age, and the major occupation recode (U.S. Census Bureau 2002:76).[14] As Hirsch and Schumacher (2004) discuss, a downward bias is introduced when a researcher analyzes the wage effect of a variable that was not included in the matching process. In our case, because the Census uses the major occupation recode (a set of 14 occupation categories), the imputation process does not match workers at the detailed occupation level within their major occupation category. It is possible, for example, that the missing wages of a nuclear engineer are imputed using the wages of a podiatrist—provided they had the same race, gender, age, and education—because both occupations are classified under the major occupation recode as "professional." Consequently, using the imputed wages converts some of the true between-occupation inequality into within-occupation inequality.

As noted earlier, Kim and Sakamoto (2008) use the Census Bureau's hot-deck procedure in their analysis of trends in within- and between-occupation inequality. Because the proportion of workers with imputed wages has increased over time (Table 2; see also Bollinger and Hirsch 2005; Hirsch and Schumacher 2004), the problem of misallocating imputed wages among detailed occupations is becoming more severe. This suggests that at least some—and perhaps all—of the increase in within-occupation inequality that Kim and Sakamoto report may be due to the bias introduced by the hot-deck imputation procedure.

To gauge the impact the Census Bureau hot-deck approach has on the trend in occupational inequality, we replicate the Census approach twice, once using the big occupational categories of the Census approach, and then using the detailed 3-digit occupations of our analysis. To ensure a large enough sample size within each of the detailed occupation categories, we use a simplified list of other variables for the match: education (college/no-college), age (under 30, 30 to 50, over 50), race (white/non-white), and gender. We sequentially match cases with missing wage data to cases with the same sets of values on the matching variables.[15] By calculating time trends in between- and within-occupational inequality for both of these hot-deck approaches, we test whether we can reproduce the Census hot-deck results with our big-occupation hot-deck. We then compare this with the imputations based on the detailed occupation hot-deck.

Multiple imputation. Our third approach uses multiple imputation of missing wages based on our individual-level model in Equations 3a and 3d, using coefficient estimates of the controls and the occupational fixed effects. For each case with missing data, we calculate the predicted wage and variance and then generate five imputed cases, each with one-fifth the sampling weight of the original case. Each imputed wage consists of the predicted mean plus a random draw from a normal variable with the predicted variance (for a discussion of multiple imputation, see Allison 2001). Multiple imputation with random draws from the estimated variance is crucial for this approach, because imputing missing wages to the predicted mean would lead to an underestimation of within-occupation variance. For our descriptive results (Table 3), we impute wages for each year of data using models of the mean (Equation 3a) and variance (Equation 3d) based on occupation, education, age, race, gender, and union membership. For our multivariate results (Tables 5 through 7), we use the same variables but combine two years of CPS data to increase precision and impute separately by gender. Table 4 provides an example of the results for mean and variance for the multivariate results.[16]

No imputation, adjusted weights. Finally, for comparison purposes we estimate trends in occupational inequality excluding the wages of imputed cases, but including all cases in the calculation of the sizes of different occupations. Although most studies using CPS data routinely exclude all cases with missing data (e.g., Autor et al. 2006; Lemieux 2006a; Weeden et al. 2007), it is possible that some occupations are more likely than others to have missing data. Deleting missing cases would then underrepresent these occupations. In this approach, we calculate the trend in inequality using Equation 2 with the average wage, μ_{jt}, and variance for each occupation, σ^2_{jt}, using the non-missing data, but we calculate the size of each occupation, p^{ALL}_{jt}, using all the cases:

$$Var[\ln wage|occ] = \sum_j p^{ALL}_{jt}(\mu_{jt} - \bar{\mu}_t)^2 + \sum_j p^{ALL}_{jt}\sigma^2_{jt} \tag{6}$$

If the evidence in Column 7 of Table 2 is correct, and no time trend exists in the difference between the variance of wages for missing and non-missing cases, then this adjusted weights approach should be the most transparent way to estimate descriptive trends in occupational inequality.

RESULTS

Trends in Within- and Between-Occupation Inequality

Table 3 breaks the trend in the variance of log wages, our measure of inequality, into within- and between-occupation components using Equation 2 and presents the results based on multiple imputation and the Census hot-deck approach. Figure 2 shows results based on the rest of our competing approaches for dealing with missing wage data.

The key finding from Table 3 is the discrepancy in the trends in between-occupation inequality using multiple imputations (MI) and the Census hot-deck imputations. We first focus on the time series using the consistent 1980 occupation codes between 1983 and 2002. Using the MI series, between-occupation inequality increases by .022 (.129 –

.107), while it increases by .009 (.112 – .103) in the hot-deck series. As an overall summary measure of the explanatory power of occupations, we calculate occupational R^2 as the ratio of between-occupation inequality to total inequality. For the MI approach, this starts at .382 in 1983, goes up to .397 in 1988, and is .387 in 2002, the last year of the 1980 occupation codes. For the Census hot-deck approach, by contrast, the R^2 declines from .363 in 1983 to .339 in 2002.

As discussed earlier, the problem with the hot-deck approach is that it imputes missing wages using broad occupational categories rather than detailed occupations; this can convert between-occupation inequality into within-occupation inequality. The severity of this bias has increased because the proportion of workers whose wages are imputed has grown from .143 in 1983 to .337 in 2002. The MI approach, by contrast, imputes wages within detailed 3-digit occupations, so it is not vulnerable to this bias.

We now turn to our attempt to extend the time series on inequality to 2008 using the 2000 occupation codes. As discussed in the Data section, the Bureau of Labor Statistics re-released the CPS data from 2000 to 2002 with occupations double-coded in both the 1980 and 2000 codes, to allow researchers to explore the differences between the two sets of occupational classifications. Comparing between- and within-occupation inequality from 2000 to 2002 for both sets of codes allows us to piece together these two time series: the average gap between the two estimates during the overlap period (2000 to 2002) suggests the degree to which within- and between-inequality should be adjusted when comparing results based on the 1980 versus 2000 occupation codes.

Using the 2000 codes, the recent trend in occupational inequality clearly shows no evidence of a decrease in between-occupation inequality up to 2008 (see Figure 2 and Table 3). Using the MI series in Table 3 with 2000 occupation codes, between-occupation inequality is .132 in 2000, .137 in 2003, and .142 in 2008. The occupational R^2 of the MI series is .433 in 2008 compared with .382 in 1983. The importance of occupations in explaining inequality has clearly gone up over time, partly due to the change in occupational coding schemes. This change was not arbitrary but was an attempt to keep up with a constantly changing division of labor (e.g., the 2000 codes contain codes for several computer and network-related occupations that were not in the 1980 codes). The increase in R^2 between the two sets of codes is thus a consequence of an improved mapping of Census occupational codes onto what people actually do.

Table 3 also shows the decomposition of changes in between-occupation inequality into components based on wage and size changes, as depicted in Equation 2b. This is interesting because two sources could cause occupational polarization: (1) wage changes, as high- and low-wage occupations spread farther apart, or (2) size changes, through an increase in the relative size of high- and low-wage occupations and a hollowing out of the middle. Table 3 shows that the majority of the increase in between-occupation inequality since 1983 is due to changes in the average wages of occupations, not the size of occupations. Between 1983 and 2008, the change in between-occupation inequality attributable to wage changes is .0243, compared with .0126 for size changes.

In Figure 2, we graph time trends in the occupational R^2s for each of our alternative approaches for handling the CPS missing data and compare them with multiple imputation (dashed lines illustrate the MI approach). Results for 1980 and 2000 codes are presented in the same graph, with a period of overlap between 2000 and 2002. Panel 1 compares the Census hot-deck with MI, which reproduces the results reported in Table 3. The gap between the two series indicates the lower R^2 for the hot-deck approach, and it is clear that the gap increases with the 2000 occupation codes. Panel 2 presents results for our big occupation hot-deck. Visually, this approach appears very similar to the Census hot-deck in Panel 1. Compared with MI, there is a decline in R^2 over time as the proportion of imputed cases increases (see Table 2).

Panel 3 shows results for the detailed occupation hot-deck imputation. In contrast to Panels 1 and 2, here the results closely mirror those of MI, such that it is difficult to see the dashed line of the MI results. Comparing the results for the detailed occupation imputation versus the big occupation imputation (Panel 2), which uses the same matching variables except for occupation, indicates that the deterioration in R^2 in the big occupation approach is the result of matching on broad occupational categories rather than on detailed occupations. Panel 3 indicates there is nothing wrong with the hot-deck approach per se, but it should include matching on detailed occupations if that is going to be the key independent variable of interest. Hirsch and Schumacher (2004) obtain similar results when they reproduce the Census hot-deck approach including matching on their key independent variable, union membership. Finally, Panel 4 shows the no imputation approach, which excludes the wages of missing cases but counts them when calculating occupational sizes (see Equation 6). The results for this approach closely match those for MI.

Kim and Sakamoto (2008) conclude that the ability of occupations to explain inequality is declining. However, results from Column 7 of Table 2, Table 3, and Figure 2 indicate that their conclusion is misleading. In particular, the difference between Panels 1–2 and 3 of Figure 2 indicates that the problem is hot-deck imputation using broad occupation groups. By contrast, hot-deck imputation using detailed occupations (Panel 3) closely matches results using multiple imputations. It appears that much of the increase in within-occupation inequality is an artifact of the Census hot-deck procedure, which corresponds to the carefully documented case of hot-deck-induced bias discussed by Hirsch and Schumacher (2004). Using either multiple imputation, detailed occupational hot-deck imputation, or excluding imputed wages, we find the opposite result of Kim and Sakamoto (2008): Table 3 and Figure 2 show that the explanatory power of occupations did not decrease from 1983 to 2002 (using the 1980 occupation codes), and it increased from 2000 to 2008 (using the 2000 occupation codes).

Individual-Level Models: Education and Occupation Fixed Effects

Having completed a reanalysis of the descriptive trend in occupational inequality, we now turn to our multivariate results, where we consider three competing explanations of why inequality increased in the United States over the past 25 years. While the between-occupation perspective argues that changes in occupational structure are the root cause, it is possible that increases in education and experience also explain some of the expansion of inequality. Not only do average wages increase with education and labor market experience, but so does the variance of wages (Lemieux 2006a). By contrast, the residual perspective argues that the unexplained component of inequality has increased over time, whether that is the wage residual net of occupation effects (Kim and Sakamoto 2008) or controlling for the returns to observed levels of education and demographic variables (Acemoglu 2002).

We can resolve these competing explanations by testing a model that simultaneously incorporates the effects of occupational structure, education (and other aspects of demographic composition), and residual inequality. Until now, studies arguing for the importance of occupations in explaining inequality (e.g., Autor et al. 2006; Weeden et al. 2007) have been limited because they do not include individual-level variables that also affect the mean and variance of wages. At the same

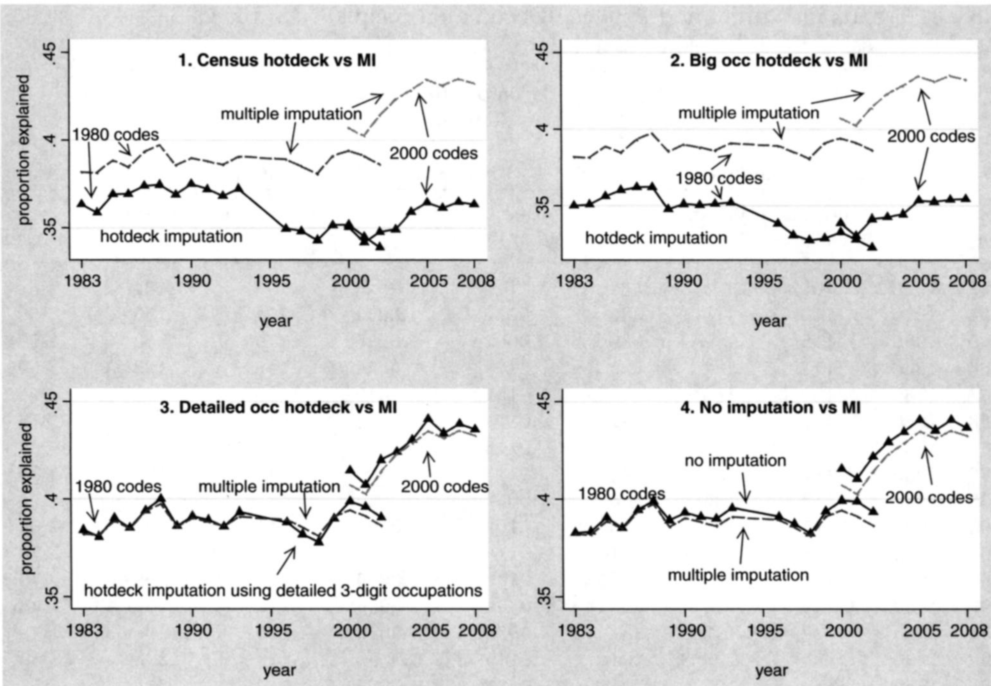

Figure 2. Comparison of Alternative Approaches to Missing Data: Proportion of Inequality Explained by Occupation Codes (occupation R^2)
Note: Dashed lines in each graph represent multiple imputation (MI).

time, the most glaring omission is the absence of occupation-level data in empirical tests of trends in residual inequality (e.g., Acemoglu 2002; Katz and Autor 1999).

We now estimate trends in between- and within-occupational inequality net of individual factors. As described earlier, we first estimate fixed-effects models of the mean and variance of wages. We then recover the occupation fixed effects using Equation 3c and use the fixed effects to calculate net occupational inequality using Equation 4. We group the data together by two-year periods to increase the sample size, and we run the models separately for men and women. We use multiple imputations to impute the wages of missing cases, adding five replicates for each missing case, each with one-fifth the sampling weight of the original.

The first step is to estimate the fixed-effects models. We estimate three different models for each two-year period by gender.[17] The first model includes only the occupation fixed effects. It is essentially the same as the descriptive results in Table 3, with the exception of the disaggregation by gender and the combination of years (in contrast to Table 3, missing data is imputed separately by gender). The second model adds age, race/ethnicity, and labor union membership as control variables. The third model adds dummy variables for education categories in addition to the demographic controls of Model 2. As an example of the first-step estimates, Table 4 shows the fixed-effects results for Model 3 for 1992/1993 and 2006/2007 for men.[18] We estimate identical models for all the two-year combinations shown in Table 5. We code all of the education and control variables as dummy variables. In 1992, the CPS switched to a credential-based system of coding education, rather than number of

Table 3. Trends in Within- and Between-Occupation Inequality in the CPS, 1983 to 2008, Based on Missing Data Procedure Used

		Multiple Imputation			Decomposition of Δ in Between-Occ. Inequality Since 1983[a]		Census Hot-Deck Imputations		
Year	Occ. Codes	Between Occs.	Within Occs.	Occ. R^2	Due to ΔAvg Wage	Due to ΔSize	Between Occs.	Within Occs.	Occ. R^2
1983	1980	.107	.174	.382	0	0	.103	.181	.363
1984	1980	.110	.178	.382	.0037	−.0011	.104	.186	.359
1985	1980	.115	.181	.389	.0092	−.0016	.110	.188	.369
1986	1980	.117	.187	.385	.0112	−.0014	.113	.193	.369
1987	1980	.121	.186	.394	.0152	−.0016	.116	.194	.374
1988	1980	.123	.187	.397	.0173	−.0014	.117	.196	.375
1989	1980	.118	.188	.386	.0116	−.0004	.112	.192	.369
1990	1980	.119	.186	.390	.0112	.0008	.114	.190	.375
1991	1980	.118	.185	.388	.0086	.0018	.112	.188	.372
1992	1980	.118	.187	.386	.0102	.0003	.111	.191	.368
1993	1980	.121	.188	.391	.0114	.0020	.114	.191	.373
1996	1980	.124	.195	.389	.0132	.0046	.112	.208	.349
1997	1980	.123	.196	.385	.0109	.0056	.112	.209	.348
1998	1980	.122	.197	.381	.0078	.0075	.109	.209	.343
1999	1980	.125	.195	.391	.0115	.0083	.112	.207	.351
2000	1980	.128	.197	.394	.0131	.0097	.114	.211	.352
2001	1980	.128	.200	.391	.0132	.0102	.112	.214	.345
2002	1980	.129	.204	.387	.0121	.0117	.112	.219	.339
2000	2000	.132	.192	.407	.0194	.0068	.114	.212	.350
2001	2000	.132	.197	.402	.0188	.0078	.112	.215	.342
2002	2000	.138	.195	.414	.0221	.0100	.116	.217	.348
2003	2000	.137	.187	.423	.0220	.0087	.111	.207	.349
2004	2000	.139	.186	.429	.0242	.0088	.115	.206	.359
2005	2000	.143	.186	.435	.0274	.0095	.119	.207	.364
2006	2000	.142	.187	.431	.0265	.0094	.117	.208	.361
2007	2000	.142	.185	.435	.0264	.0098	.119	.208	.365
2008	2000	.142	.187	.433	.0243	.0126	.119	.208	.365

[a]The decomposition of the change in between-occupation inequality (using the multiple imputation approach) from 1983 to the current year that is due to changes in the average wages of occupations or changes in the size of occupations (see Equation 2b).

years of school attended, so models prior to 1992 include dummy variables for years of education rather than the specific credential obtained. We collapse all high school drop-outs into a single category for both education codes (less than high school). The key difference between the results in Table 4 and a typical human capital wage regression is that all the variables are transformed using Equation 3b by subtracting the occupation-specific average. As a result, the coefficients in Table 4 are average within-occupation effects net of the fixed effects of occupations. Overall, results in Table 4 confirm previous research about the effect of education and age on wages. While the effects on average wages are well known, the effect on the variance is noteworthy. The variance of log wages tends to go up with increases in education, and it tends to go up over the life course in both 1992/1993 and 2006/2007.

Table 5 shows trends in the level of occupational inequality for each of our three models. Again, after we estimate the mean and variance

Table 4. Occupation Fixed-Effects OLS Results for Men in 1992 to 1993 and 2006 to 2007

	Log Wages		Variance of Log Wages	
	1992 to 1993	2006 to 2007	1992 to 1993	2006 to 2007
Education[a]				
Less than high school	−.113***	−.114***	−.00481*	−.0112***
Some college	.0511***	.0569***	.00862***	.00199
Associates (vocational)	.0643***	.0728***	.00965**	−.0069*
Associates (academic)	.0984***	.103***	.0173***	−.000706
College	.228***	.225***	.0306***	.0176***
Master's degree	.353***	.351***	.0342***	.0207***
Professional degree	.424***	.386***	.0712***	.0228***
Doctorate degree	.498***	.490***	.0235***	.00415
Race[b]				
Black	−.122***	−.102***	.00781***	.00472**
Hispanic	−.0498***	−.0499***	.0216***	.0124***
Other	−.0718***	−.0911***	.00837***	.00716***
Age[c]				
20 to 24	.113***	.0991***	.0250***	.0199***
25 to 29	.263***	.210***	.0365***	.0184***
30 to 34	.369***	.312***	.0485***	.0289***
35 to 39	.429***	.377***	.0545***	.0331***
40 to 44	.468***	.401***	.0584***	.0418***
45 to 49	.515***	.419***	.0706***	.0389***
50 to 54	.517***	.407***	.0763***	.0469***
55 to 59	.481***	.392***	.0857***	.0459***
60 to 64	.457***	.364***	.0930***	.0652***
Union	−.232***	−.237***	.0198***	.0110***
Constant	4.97e-10	−2.67e-10	−2.43e-11	−4.80e-11
N[d]	188,964	366,859	188,962	366,856

Note: Identical models are estimated for men and women for all the pairs of years in Table 5.
[a]Omitted category for education: high school degree.
[b]Omitted category for race: white.
[c]Omitted category for age: under 20 years.
[d]Includes five replicate cases for each respondent with imputed wages (with one-fifth the sampling weight), see Methods section.
* $p < .05$; ** $p < .01$; *** $p < .001$.

occupation fixed effects by gender using Equation 3c (as illustrated by the results for two time periods in Table 4), we estimate between- and within-occupation inequality using Equation 4. We then combine the gender-specific fixed effects using the relative proportion of men and women in each occupation (see the Methods section). The advantage of estimating separate occupation effects by gender is a precise estimate of the gender wage gap within each occupation. A within-occupation gender wage gap contributes to the overall level of within-occupation inequality (as the variance of average occupational wages by gender). Table 5 reports this within-occupation gender inequality for all years. Although it declines over time in all models, gender inequality contributes little to the overall level of within-occupation inequality, indicating that most within-occupation inequality occurs within gender and most gender inequality occurs across occupations.

In all three models in Table 5, we calculate the impact of switching from the 1980 to 2000 occupation codes by estimating each model twice on the combined data from 2000 to 2002: once using the 1980 occupation codes and once using the 2000

Table 5. Trends in Occupational Inequality, Controlling for Individual-Level Variables, Hetereoskedastic OLS Models[a]

Model 1. Baseline Model (occupation dummy variables only)

Year	Occupation Codes	Between-Occupation	Within-Occupation	Total[c]	Individual-Level Factors	Within-Occupation Gender[d]
1983/1984	1980	.1097	.1774	.2871	0	.0106
1990/1991	1980	.1186	.1869	.3055	0	.0088
1992/1993	1980	.1190	.1881	.3071	0	.0078
2000/2002	1980	.1281	.2017	.3298	0	.0071
2000/2002	2000	.1336	.1962	.3298	0	.0062
ΔDue to switch in occ. codes[b]		*+.0055*	*−.0055*	*0*	*0*	
2004/2005	2000	.1411	.1867	.3278	0	.0051
2006/2007	2000	.1419	.1877	.3296	0	.0049
2007/2008	2000	.1420	.1871	.3291	0	.0048

Model 2. Baseline + Demographic Variables (age, race, and union status)

Year	Occupation Codes	Between-Occupation	Within-Occupation	Total[c]	Individual-Level Factors	Within-Occupation Gender[d]
1983/1984	1980	.0864	.1500	.2364	.0507	.0111
1990/1991	1980	.0921	.1576	.2497	.0559	.0078
1992/1993	1980	.0921	.1564	.2485	.0586	.0068
2000/2002	1980	.1028	.1713	.2740	.0558	.0063
2000/2002	2000	.1077	.1670	.2746	.0551	.0056
ΔDue to switch in occ. codes[b]		*+.0049*	*−.0043*	*+.006*	*−.007*	
2004/2005	2000	.1118	.1582	.2700	.0578	.0050
2006/2007	2000	.1131	.1591	.2722	.0574	.0049
2007/2008	2000	.1142	.1584	.2726	.0565	.0047

Model 3. Model 2 + Education

Year	Occupation Codes	Between-Occupation	Within-Occupation	Total[c]	Individual-Level Factors	Within-Occupation Gender[d]
1983/1984	1980	.0655	.1313	.1968	.0903	.0096
1990/1991	1980	.0629	.1373	.2001	.1054	.0065
1992/1993	1980	.0600	.1321	.1921	.1150	.0054
2000/2002	1980	.0644	.1397	.2041	.1257	.0052
2000/2002	2000	.0694	.1401	.2095	.1202	.0052
ΔDue to switch in occ. codes[b]		*+.0050*	*+.0004*	*+.0054*	*−.0055*	
2004/2005	2000	.0737	.1357	.2094	.1184	.0048
2006/2007	2000	.0733	.1374	.2107	.1188	.0047
2007/2008	2000	.0746	.1397	.2142	.1149	.0048

[a]See Equation 4 for derivation.
[b]Difference between 2000 to 2002 with 2000 occupation codes and 1980 occupation codes (see text).
[c]"Total" is the sum of the between- and within-occupation components.
[d]"Within-Occupation Gender" refers to the component of within-occupation inequality that is due to the within-occupation gender wage gap (see text).

occupation codes. For each model, the row labeled "Δ Due to switch in occ. codes" shows the change in between-, within-, and individual-level inequality when we go from the 1980 to 2000 occupation codes for the years 2000 to 2002.

The results for Model 1 in Table 5 use occupation fixed effects with no controls. These results confirm the basic time trends discussed in Table 3. Dividing the data into two periods, one from 1983/1984 to 1990/1991 and the other from 1992/1993 to 2007/2008, almost all of the increase in within-occupation inequality is concentrated in the first period. By contrast, the growth in between-occupation inequality is concentrated in the second period (from .1190 in 1992/1993 to .1420 in 2007/2008). However, according to the results with the double-coded occupation data in 2000 to 2002, about .0055 of that increase is due to the change from the 1980 to 2000 occupation codes: between-occupation inequality jumps from .1281 to .1336 when we switch to the 2000 occupation codes. One could simply adjust the change in between-inequality downward by this amount, .0055, to reflect the component attributable to the change in occupational classification systems. On the other hand, the higher level of between-occupation inequality in the 2000 codes indicates their greater accuracy in describing what people actually do. During this period, the actual occupation structure changed continuously, rendering the classification system outdated. As new occupations emerge, using an antiquated occupational classification system will result in an apparent increase in within-occupation inequality. Identifying the effect of the change in coding on the results in Table 5 makes transparent the degree to which the new occupational codes affect the results.

What is the effect on these trends after we control for individual-level variables that also affect inequality? Model 2 in Table 5 presents trends in occupation-level inequality controlling for demographic variables. This includes all of the variables in Table 4 except education. The level of inequality explained by occupations for each year has dropped. For example, although the level of between-occupation inequality in 1983/1984 was .1097 using Model 1, it declined to .0864 for the same period in Model 2. The reason for this difference, of course, is that some part of the average wage of each occupation is explained away as the result of occupation-level differences in observed covariates in Equation 3c. As a result, the inequality generated by the occupation-level fixed effects is smaller using the results from Model 2 than those from Model 1. The level of within-occupation inequality also decreases, as some of the variance in wages within occupations is attributed to individual-level factors such as age, ethnicity, and union membership. As discussed earlier, the impact of age on within-occupation inequality should not be surprising. Results in Table 4 indicate that the within-occupation variance of log wages increases as age increases.

Model 3 in Table 5 adds education, demographic variables, and occupation fixed effects. Here we see that between-occupation inequality actually declines from 1983/1984 to 1990/1991, from .0655 to .0600, but increases substantially afterward to .0746 in 2007/2008. By contrast, the results for within-occupation inequality confirm results presented in Model 1: most of the increase in within-occupation inequality is between 1983/1984 and 1990/1991. Notably, some of the decline in within-occupation inequality between 1990/1991 and 1992/1993 may be due to the shift in coding of education from years attended to credentials. The credential approach may have slightly more explanatory power. As a result, it is convenient to base our discussion of time trends on the periods 1983 to 1991 and 1992 to 2008.

Although the level of inequality attributable to occupation-level effects (between or within) declines once we add individual-level covariates in Models 2 and 3, what we really care about are trends over time. To make the time trends more transparent, Table 6 presents a decomposition of changes in inequality over time based on each model. The first column is the change in the actual level of inequality, taken from Model 1 in Table 5. This was .0185 from 1983/1984 to 1990/

Table 6. Decomposition of Changes in Inequality, 1983 to 1991 and 1992 to 2008

	Model 1: Baseline	Model 2: Age, Race, Union	Model 3: M2 + Education	Percent of Change, in Model 3
1983 to 1991				
Overall (from Model 1)	.0185			
ΔBetween	.0089	.0057	−.0026	−14%
ΔWithin	.0095	.0076	.0060	32%
ΔIndividual-Level Variables	0	.0052	.0151	82%
1992 to 2008				
Overall (from Model 1)	.0220			
ΔBetween	.0231	.0221	.0146	66%
(ΔBetween due to change in occ. codes)	(.0055)	(.0049)	(.0050)	(23%)
ΔWithin	−.0011	.0020	.0076	35%
(ΔWithin due to change in occ. codes)	(–.0055)	(–.0043)	(.0004)	(2%)
ΔIndividual-Level Variables	0	−.0021	−.0001	0%
(ΔInd. due to change in occ. codes)	0	(–.0007)	(–.0055)	(–25%)

1991 and .0222 from 1992/1993 to 2007/2008. The change in between- and within-occupation inequality is the difference between the two time points for each measure. The change attributed to individual-level variables is the change in inequality not explained by changes in between- and within-occupational inequality. In both periods, the impact of individual-level variables in Model 2 is due to demographic factors, while the change in the level attributed to individual variables between Models 2 and 3 illustrates the impact of adding education to the model. For the period from 1992 to 2008, we include, in parentheses, the change in each component attributable to the shift from the 1980 to the 2000 occupation codes, as calculated in Table 5 by the difference between the results for the double-coded data from 2000 to 2002 using each set of codes.

In the first period, 1983/1984 to 1990/1991, Model 3 shows that the increase in within-occupation inequality does explain a substantial proportion of the increase (.006 of the .0185 increase), but not as much as the amount explained by the individual-level variables (.0151). In the second period, 1992/1993 to 2007/2008, however, between-occupation inequality is dominant: .0146 out of an overall increase of .0220, compared with .0076 for within-occupation inequality and –.0001 for individual-level effects. Based on results in Model 3 for the second period, we attribute 66 percent of the change in inequality from 1992/1993 to 2007/2008 to between-occupation fixed effects. This includes a .0050 increase (23 percent) from the switch in occupational codes (as calculated using the double-coded 2000 to 2002 data). In contrast to the effect for between-occupation inequality, we attribute 35 percent to within-occupation inequality, and essentially 0 (–.2 percent) to individual-level variables, including education. This does not mean that education and demographic variables do not matter for inequality from 1992 to 2008. Indeed, in Model 3 of Table 5 these factors explain a substantial .1149 of the overall level of inequality (.3291) in 2007/2008. It just means that the extent of their effect has not changed since 1991/1992, when it was .1150. By contrast, the net effect of between- and within-occupation inequality has increased. Overall, the results in Tables 5 and 6 show substantial evidence of occupation-level effects on inequality during the past 15 years, net of demographic and education factors at the individual level that also contribute to increases in inequality.

Sensitivity Analysis: Results for Individual Occupations

One of the benefits of our approach is that, by calculating mean and variance occupation effects, we can explore the impact of individual occupations on inequality over time. It is tempting to interpret the finding of increasing between-occupation inequality as resulting from very broad and sweeping changes in the occupation structure, such as a growth in "good" and "bad" jobs. Using our data, we can test the accuracy of this interpretation. We use Equation 5 to estimate the change in inequality attributable to specific occupations by calculating the difference between actual inequality and the counterfactual level of inequality that would exist had occupation j held constant over time. We use the consistent set of 1980 occupation codes for the years 1983 to 2002 for this analysis.

Table 7 presents data on the 20 occupations that had the largest impact on inequality between 1983/1984 and 2000/2002. The first column of results is the change in adjusted varlog ($\times$ 1,000) based on occupation size p and mean and variance occupational fixed effects derived from results in Model 3. The second column uses the actual p, μ, and σ^2 to calculate the unadjusted change in varlog. In addition to the overall effect on inequality, Columns 3 through 5 show the change in adjusted varlog attributable to size, mean, and variance changes for each occupation using the occupation mean and variance fixed effects calculated using Equations 3a through 3d. As with the overall effect, these are calculated using Equation 5, except only one parameter for each occupation (either p, or the adjusted μ or σ^2) is held constant at a time. Note that these effects are not additive; the size of the occupation, p, is multiplied by both μ and σ^2 in Equation 2.

Columns 6 through 9 of Table 7 present descriptive statistics for these 20 occupations. Column 6 shows the adjusted mean-centered average wage for each occupation in 1983/1984, after controlling for education and demographic characteristics. This shows which occupations pay above or below the population average wage, net of observed individual characteristics. Column 7 shows the change in the size of the occupation ($\times$ 100) from 1983 to 2002. Columns 8 and 9 show the change in adjusted average wage and adjusted within-occupation variance of wages during this same period.

Based on the adjusted change in inequality using the occupation fixed effects (Column 1), managers not elsewhere classified (NEC), secretaries, and computer systems analysts had the biggest impact on inequality during this time period. In Columns 3, 4, and 5, which identify the change in adjusted varlog due to each occupation's p, μ, and σ^2, we see that size changes (Column 3) are the primary reason why these three occupations had a large impact on overall inequality. Column 7 shows that, in the case of secretaries, the impact was due to a decline in size of a middle-wage occupation (adjusted wage in 1983/1984 is –.09 in Column 6), while managers NEC and computer systems analysts were expanding occupations with above average adjusted wages in 1983/1984.

Some occupations that appear to have a big impact based on the change in unadjusted varlog (Column 2 in Table 7) have their effect largely explained away by individual covariates. For example, the adjusted change in varlog for cashiers is .52 based on occupation fixed effects, but the unadjusted change in varlog is more than five times larger, 2.60. This indicates that overall trends in returns to education or demographic variables explain most of cashiers' impact on inequality.

Using disaggregated occupations, we can evaluate the sensitivity of the trend in occupational polarization to the inclusion or exclusion of a few occupations. With

Table 7. Top 20 Occupations with the Largest Effect on the Change in Inequality, 1983/1984 to 2000/2002

	Overall Effect		Δ in Adj. Varlog Due To:				Δ from 1983/1984 to 2000/2002		
Column #	1	2	3	4	5	6	7	8	9
Occupation	ΔAdj. Varlog × 1,000	ΔVarlog × 1,000	P	μ	σ^2	1983/1984 adj.$_\mu$	ΔP	adj$_\mu$	Δadj. σ^2
Managers and admin, NEC	2.85	3.82	2.80	.25	−.19	.27	2.11	−.02	.00
Secretaries	2.36	5.08	1.55	.11	.27	−.09	−2.28	−.02	.01
Computer systems analysts	2.30	3.07	2.16	.19	.44	.45	1.23	−.02	.03
Physicians[a]	1.49	3.47	.86	.25	.71	.14	.20	.39	.12
Childcare workers, private household	−1.45	−1.57	−.17	−.79	.03	−.72	−.08	.40	.03
Securities and financial sales occs.	1.11	1.10	.93	.04	.31	.49	.21	−.03	.07
Farm workers	−.80	−1.42	−.54	.04	−.17	−.46	−.53	.01	−.03
Textile sewing machine operators	−.75	.16	−.51	−.11	.02	−.47	−.59	−.07	.01
Janitors and cleaners	.74	1.56	−.03	.73	−.06	−.29	−.24	−.10	.00
Investigators and adjusters	−.67	−.88	−.60	.04	−.26	.00	.61	−.17	−.03
Registered nurses	.64	1.79	−.18	.59	.38	.34	.27	.11	.02
Supervisors production occs.	−.59	−.21	.28	−.41	−.13	.23	−.62	−.15	−.01
Cashiers	.52	2.60	.01	.65	−.14	−.43	.02	−.12	−.01
Supervisors and proprietors	.52	.60	.09	−.02	.59	.06	.88	−.01	.02
Computer programmers	.49	.88	.05	.32	.15	.41	.04	.10	.03
Private household cleaner	−.47	−.35	−.20	−.24	.05	−.55	−.09	.06	.03
Managers, marketing and advertising	.46	.56	.46	−.14	.15	.44	.25	−.04	.02
Computer operators	.46	.93	.36	.01	.03	.04	−.48	.03	.01
Welfare service aides	.46	.77	.44	.01	.08	−.51	.30	−.11	.02
Sales workers, other commodities	−.44	−.23	−.18	−.63	.43	−.33	−.28	.13	.05

[a]CPS data does not record wages for self-employed workers, so doctors and lawyers include only those who are not self-employed.

Key:

Col. #	Variable
1	Adjusted Δ varlog: change in varlog (×1,000) due to this occupation, after controlling for education and demographic factors (see Model 3 of Table 4)
2	Δ varlog: change in varlog (×1,000) due to changes in this occupation
3–5	Change in adjusted varlog (×1,000) due to size, average wage, and within-occupation inequality (calculated using Equation 5, see text)
6	1983/1984 adj μ: adjusted (mean centered) log wage in 1983/1984 after controlling for individual-level factors (see Equations 3a and 3d)
7	ΔP: change in employment share (×100) from 1983/1984 to 2000/2002
8	Δadj$_\mu$: change in adjusted average log wage (mean centered) from 1983/1984 to 2000/2002
9	Δadj.σ^2: change in adjusted within-occupation inequality, from 1983/1984 to 2000/2002

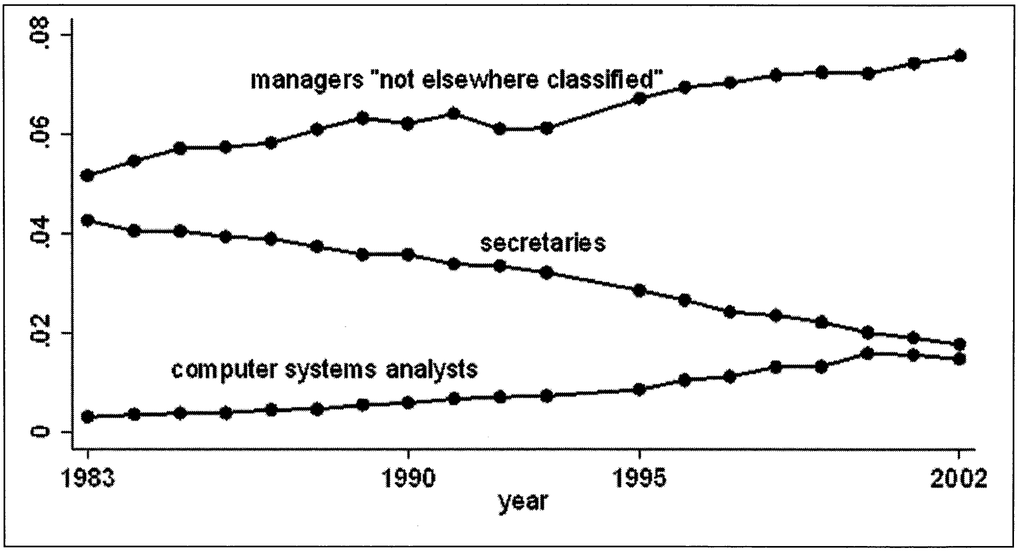

Figure 3. Change in Employment Share, 1983 to 2002
Note: Weighted by hours worked.

Equation 5, we can calculate the impact of the top three occupations—managers NEC, secretaries, and computer systems analysts—on changes in overall inequality by estimating the counterfactual change in inequality for all three occupations at the same time. From 1983/1984 to 2000/2002, there was an overall increase in inequality of .0427, based on Model 1 in Table 5. Using Equation 5, we find that .0077 (18 percent) of this overall increase in inequality is attributable to the combined effect of managers NEC, secretaries, and computer systems analysts, controlling for education and demographic variables. If we use the unadjusted varlog (Column 2), the combined effect is .0123 (29 percent) of the overall increase in inequality. In addition, the overall impact on the adjusted varlog of the top 10 occupations in Table 7 with the largest positive effect on inequality is .0136 (32 percent) of the overall increase in inequality from 1983 to 2002.[19] This result indicates that trends in occupational inequality are sensitive to the impact of a relatively small number of occupations.

Figure 3 shows trends in the employment shares of the top three inequality generating occupations over time. It is clear that the size of each occupation has been gradually changing from 1983 to 2002. Do the occupation codes mean the same thing over time, or do they mask substantial changes in the division of labor within these occupations? Occupational structure is always changing and classification of occupations depends on a particular classification scheme. This has important consequences as new occupations emerge. When several occupations are grouped together because of an antiquated occupation classification scheme, what appears to be growing within-occupation inequality may actually mask growing heterogeneity among different occupations that are combined in the same code. In particular, the expansion of employment in managers NEC and computer systems analysts deserves further discussion because these two occupations (as coded in 1980) are actually a collection of different occupations (as seen in the 2000 classification system). Table 8 uses the double-coded occupation data in the CPS from 2000 to 2002 to depict this reclassification from the 1980 to 2000 codes for these two

Table 8. Splitting of Managers "Not Elsewhere Classified" (NEC) and Computer Systems Analysts, 1980 Occupation Codes to 2000 Occupation Codes

2000 Occupation Code	N
Panel A. Cases Coded as Managers NEC in the 1980 Codes	
Managers, All Other	5,892
Chief Executives	3,658
General and Operations Managers	2,727
Food Service Managers	2,253
Secretaries and Administrative Assistants	1,644
First-Line Supervisors/Managers of Office and Admin.	1,353
Construction Managers	1,255
Industrial Production Managers	1,024
Social and Community Service Managers	836
First-Line Supervisors/Managers of Non-retail Sale	810
Transportation, Storage, and Distribution Managers	764
First-Line Supervisors/Managers of Retail Sales Workers	764
First-Line Supervisors/Managers of Food Preparation	728
Computer and Information Systems Managers	697
Lodging Managers	454
Total	24,859
All others	7,006
Panel B. Cases Coded as Computer Systems Analysts in the 1980 Codes	
Computer Software Engineers	1,873
Computer Scientists and Systems Analysts	1,546
Network Systems and Data Communication Analysts	564
Computer Support Specialists	551
Network and Computer Systems Administrators	427
Computer Programmers	185
Database Administrators	163
Computer and Information Systems Managers	98
Management Analysts	91
Managers, All Other	65
Computer Hardware Engineers	65
Operations Research Analysts	64
Computer, Automated Teller, and Office Machine Rep.	38
Computer Operators	34
Designers	34
Total	5,798
All others	588

Source: CPS data from 2000 to 2002 with the 1980 and 2000 occupation codes.

occupations. A number of additional managerial occupations have been added, including chief executives. While most of these occupations existed in 1980, giving them a specific 3-digit code in 2000 reflects the growth in size and prominence of managerial occupations in general. As a result, Table 8 shows that most of the respondents in the 2000 to 2002 CPS data who were classified as managers NEC in the 1980 codes were reclassified into specific managerial occupations in the 2000 occupational codes. The recoding of computer systems analysts is even more striking. Here we find a number of occupations that did not exist in 1980. The upshot of Table 8 is that the increase in between-occupation inequality in 2001/2002, when we go from the 1980 to the 2000 codes, is not just a reflection of arbitrary occupational classifications. Rather, it is picking up the growing employment share and increasing division of labor in expanding fields such as managerial and computer-related occupations.

DISCUSSION AND CONCLUSIONS

Wage inequality in the United States increased dramatically in the 1980s, leveled off in the 1990s, then increased more slowly. We reexamine the role of occupational changes in explaining these patterns of wage inequality. We model the mean and variance of wages for each occupation, controlling for education and demographic factors at the individual level, to test three competing explanations of the increase in wage inequality: between-occupation polarization, changes in education and demographic composition, and within-occupation residual inequality (net of human capital variables). After correcting for a problem with imputed data that biased Kim and Sakamoto's (2008) results, we find that occupational differences in wages actually became *more* salient for explaining wage differences during the past 15 years (see Table 6). Between-occupation changes explain 66 percent of the increase in inequality from 1992/1994 to 2007/2008, and the explanatory power of occupations (the proportion of inequality explained by occupation codes) has risen from .382 in 1983 to .433 in 2008. In addition, the rise in occupation inequality appears relatively concentrated in a handful of occupations. Sensitivity analysis reveals that 18 percent of the overall increase in inequality from 1983 to 2002 results from changes in just three occupations: managers NEC, secretaries, and computer systems analysts.

These patterns are consistent with many theoretical arguments and conjectures about changes that occurred in the structure of work and occupations during the past several decades. In particular, our results support the view of a growing polarization of occupations, especially in the 1990s, as reflected in the growth of high- and low-wage occupations and the decline of middle-level occupations. High-paying managerial and professional occupations (especially lawyers, architects, engineers, computer scientists, and bankers) increased because of technological innovations, deregulation of markets, and opportunities presented by globalization. On the other hand, the expansion of the service sector accelerated the shift from manual to white-collar and service workers. This trend occurred over the twentieth century (Wyatt and Hecker 2006) and contributed to the growth of low-paying sales and service occupations in industries such as retail trade, temporary services, janitorial, home health aides and nursing home facilities, child daycare, and restaurants and food service. Middle-level occupations have also declined. Administrative support occupations (e.g., secretaries, administrative assistants, typists, and file clerks) have decreased because of organizations' downsizing and the growth of offshore white-collar technical and computer-related jobs, especially in the 1990s. Blue-collar occupations (e.g., mechanics, precision production, and operators) declined due to a combination of trade and technology (e.g., the computerization and routinization of occupations such as operatives) and the more general decrease in manufacturing workers.

A growing polarization of occupations at the top and bottom of the wage distribution—with a trough or hollowing in the middle—has important implications for issues of social mobility and careers. For example, increased polarization could mean fewer opportunities for those at the bottom to move into better-paying occupations. The decrease in middle-income occupations also contributed to the decline in the size of the middle class in the United States.

At the same time, we find that a substantial proportion of the change in wage inequality during the past quarter-century is attributable to a small number of occupations. This finding cautions us against making general, sweeping statements about the extent of polarization. The effects of the polarization of occupations may be focused on a relatively small number of occupations, rather than being a general feature of the occupational structure. This result underscores the utility of studying specific occupational labor markets that may be key to understanding the sources of the recent increase in wage inequality, rather than examining patterns of inequality for the entire labor force.

Focusing on particular occupations would also facilitate analyses that seek to integrate organizational and occupational explanations of inequality. Our finding that most of the increase in wage inequality is due to differences between occupations does not imply that diversity among firms and individuals' bargaining power within firms are not important sources of inequality. Analyzing specific occupations in depth would highlight the impact produced by variation in employees' skill, authority, and other organizational characteristics. It would help us better understand how similar kinds of occupational activities may be organized differently among workplaces depending on companies' strategies or competitive pressures produced by product markets (Baron and Bielby 1980; Lindbeck and Snower 1996).

Unfortunately, the utility of the CPS data are limited for assessing the mechanisms generating within-occupation inequality because they do not include information on organizational characteristics. Moreover, the CPS data provide information only on cross-sections of the labor force and do not permit the assessment of changes in inequality associated with a career or life course. Attempts to explain how wage inequality is generated need to disentangle mechanisms that are due to career life-cycle differences in wages (i.e., some people are better able than others to improve their wages over their careers) from those due to labor market segmentation (i.e., some people continue to receive relatively high or low wages throughout their careers). This requires the use of longitudinal panel data so as to examine earnings mobility or changes in earnings over time. Linked employer-employee datasets (such as the Census Bureau's Longitudinal Employer-Household Dynamics surveys) are also needed to account for differences in the experiences and earnings trajectories of employees in different organizations.

Sociologists are well-positioned to build theories of occupational change based on political, institutional, and organizational forces that shape the structure of work and the economy. Understanding the ways in which occupational differences generate wage inequality is a key issue for assessing how economic restructuring affects social stratification, and it is central to theories and policies related to labor market segmentation. The links between occupations and inequality are also a fertile area for integrating economic and sociological approaches to studying labor markets and inequality.

Authors' Note

A version of this article was presented at the August 2006 meeting of the American Sociological Association, Montreal, Canada.

Notes

1. Wright and Dwyer (2003:303) note that trends in income inequality are not necessarily related to job growth across job quality quintiles. Increasing income inequality could be associated with flat growth because (1) the income spread between the best and worst jobs could be increasing even if they were growing at the same rate, and (2) increasing earnings inequality can occur within job types.
2. In 2003, the CPS switched to 2000 occupation codes. We calculate p_j^{1983} for the 2000 codes using the 2000 to 2002 CPS data that is double coded with the 1980 and 2000 occupations. This calculation creates a matrix of the probability that each 1980 code is reclassified into each 2000 code, which we use to translate the 1983 occupational distribution into the 2000 codes. The computer code for these calculations is available from the first author on request.
3. To calculate the net effect of occupations controlling for individual-level variables, it might be

tempting to residualize the effect of education and demographic variables by first running wage regressions for these variables alone, predicting the residuals, and calculating occupational effects as the average residual by occupation. However, if respondents' occupations are correlated with controls such as education, then omitting occupations from the regression will bias estimates of the occupational component.

4. We estimate consistent standard errors using the Huber-White procedure.
5. We rescale the weights so that $\sum p_{it} = 1$ when estimating the counterfactual inequality.
6. This leaves the variance of log wages unaffected but has an attractive feature for the analysis: an occupation that has constant wages over time while overall wages increase contributes to an increase in inequality because it does not keep up with the general change in average wages.
7. The Stata code necessary to replicate the entire analysis is posted on the first author's Web site (http://www.unc.edu/~tedmouw/papers/papers.htm).
8. We exclude waiters and waitresses (1980 occupation code 435) from the analysis because of an unexplained jump in average (inflation adjusted) wages of $2.90 from 1988 to 1989, and $1.26 from 1993 to 1994. This likely reflects changes in recording of tips, but it is not documented in the data.
9. In 2005, the flag identifying imputed cases is present in August through December.
10. The top codes on edited weekly earnings were $999 from 1973 through 1988, $1,923 from 1989 through 1997, and $2,884 from 1998 through 2002.
11. The reasonableness of this assumption can be verified by comparing the right tail of the wage distribution between 1988 and 1989, when the top code increased from $999 to $1,923. The mean weekly earnings of workers with wages greater than $999 (the top code in 1988) in 1989 is $1,351 (this includes the .005 who were top coded at the new top code of 1.4*$1,923), close to 1.4 times the original top code of $999.
12. There is evidence that the proportion of income going to the top 1 percent of the population has risen since 1990, at least as measured by tax data (Piketty and Saez 2006). Some of this increase, however, may be due to changes in the tax code that moved certain components of business income onto individual tax returns (see Reynolds 2007).
13. In 1992, the CPS switched to the 1990 Census occupational codes, but the changes are minor and easily recoded (see the online supplement for details).
14. The categories of major occupation are (1) executive, administrative, and managerial occupations, (2) professional, (3) technical and related support, (4) sales, (5) administrative support and clerical, (6) private household, (7) protective service occupations, (8) service, (9) precision production, (10) machine operators, (11) transportation, (12) handlers and laborers, (13) farming, and (14) armed forces.
15. Approximately 1 percent of missing cases in the detailed occupation hot-deck cannot be matched within their specific matching variables. These are matched to a non-missing case in the next group of matching variables, within the same occupation. The computer code necessary to replicate this procedure is available on the first author's Web site.
16. An alternative approach would use the predicted mean and variance of each missing case in direct calculations of the level of between-occupation inequality.
17. As discussed earlier, we estimate separate models by gender because of the high degree of occupational segregation between men and women.
18. Full results for all models and years are available on the first author's Web site.
19. These 10 occupations represent 20.9 percent of the labor force in the 2000 to 2002 CPS data.

References

Acemoglu, Daron. 2002. "Technical Change, Inequality, and the Labor Market." *Journal of Economic Literature* 40:7–72.

Allison, Paul. 2001. *Missing Data*. Thousand Oaks, CA: Sage Publications.

Autor, David H., Lawrence F. Katz, and Melissa S. Kearney. 2006. "The Polarization of the U.S. Labor Market." *American Economic Review* 96:189–94.

———. 2008. "Trends in U.S. Wage Inequality: Re-Assessing the Revisionists." *Review of Economics and Statistics* 90:300–23.

Baron, James N. and William T. Bielby. 1980. "Bringing the Firms Back In: Stratification, Segmentation, and the Organization of Work." *American Sociological Review* 45:737–65.

Bernhardt, Annette, Martina Morris, Mark S. Handcock, and Marc A. Scott. 2001. *Divergent Paths: Economic Mobility in the New American Labor Market*. New York: Russell Sage Foundation.

Bielby, William T. and Arne L. Kalleberg. 1981. "The Structure of Occupational Inequality." *Quality and Quantity* 15:125–50.

Blau, Peter and Otis Dudley Duncan. 1967. *The American Occupational Structure*. New York: Wiley.

Bollen, Kenneth and Jennie E. Brand. Forthcoming. "A General Panel Model with Random and Fixed Effects: A Structural Equations Approach." *Social Forces*.

Bollinger, Christopher R. and Barry T. Hirsch. 2005. "Match Bias from Earnings Imputation in the Current Population Survey: The Case of Imperfect Matching." IZA Discussion paper #1846 (November).

Card, David and John E. DiNardo. 2002. "Skill-Biased Technological Change and Rising Wage Inequality: Some Problems and Puzzles." *Journal of Labor Economics* 20:733–83.

DiPrete, Thomas A. 2007. "What Has Sociology to Contribute to the Study of Inequality Trends? A Historical and Comparative Perspective." *American Behavioral Scientist* 50:603–18.

Featherman, David L. and Robert M. Hauser. 1978. *Opportunity and Change*. New York: Academic Press.

Goos, Maarten and Alan Manning. 2007. "Lousy and Lovely Jobs: The Rising Polarization of Work in Britain." *The Review of Economics and Statistics* 89:118–33.

Grusky, David B. and Jesper B. Sørensen. 1998. "Can Class Analysis Be Salvaged?" *American Journal of Sociology* 103:1187–234.

Hirsch, Barry T. and Edward J. Schumacher. 2004. "Match Bias in Wage Gap Estimates due to Earnings Imputation." *Journal of Labor Economics* 22:689–722.

Juhn, Chinhui, Kevin M. Murphy, and Brooks Pierce. 1993. "Wage Inequality and the Rise in Returns to Skill." *Journal of Political Economy* 101:410–42.

Katz, Lawrence F. and David H. Autor. 1999. "Changes in the Wage Structure and Earnings Inequality." Pp. 1463–1555 in *Handbook of Labor Economics*, Vol. 3A, edited by O. Ashenfelter and D. Card. Amsterdam, Netherlands: Elsevier Science.

Katz, Lawrence F. and Kevin M. Murphy. 1992. "Changes in Relative Wages, 1963–1987: Supply and Demand Factors." *The Quarterly Journal of Economics* 107:35–78.

Kim, Changhwan and Arthur Sakamoto. 2008. "The Rise of Intra-Occupational Wage Inequality in the United States, 1983 to 2002." *American Sociological Review* 73:129–57.

Lemieux, Thomas. 2006a. "Increasing Residual Wage Inequality: Composition Effects, Noisy Data, or Rising Demand for Skill?" *American Economic Review* 96:461–98.

———. 2006b. "Postsecondary Education and Increasing Wage Inequality." *American Economic Review* 96:195–99.

———. 2008. "The Changing Nature of Wage Inequality." *Journal of Population Economics* 21:21–48.

Levy, Frank and Richard J. Murnane. 1992. "U.S. Earnings Levels and Earnings Inequality: A Review of Recent Trends and Proposed Explanations." *Journal of Economic Literature* 30:1333–81.

Lindbeck, Assar and Dennis J. Snower. 1996. "Reorganization of Firms and Labor Market Inequality." *American Economic Review* 86:315–21.

Madrian, Brigitte and Lars J. Lefgren. 2000. "An Approach to Longitudinally Matching Current Population Survey (CPS) Respondents." *Journal of Economic and Social Measurement* 26:31–62.

Massey, Douglas S. and Deborah S. Hirst. 1998. "From Escalator to Hourglass: Changes in the U.S. Occupational Wage Structure 1949–1989." *Social Science Research* 27:51–71.

Mincer, Jacob. 1974. *Schooling, Experience and Earnings*. New York: Columbia University Press.

Morris, Martina and Bruce Western. 1999. "Inequality in Earnings at the Close of the Twentieth Century." *Annual Review of Sociology* 25:623–57.

Myles, John. 2003. "Where Have All the Sociologists Gone? Explaining Economic Inequality." *Canadian Journal of Sociology* 28:551–59.

Parkin, Frank. 1971. *Class Inequality and Political Order*. London, UK: Paladin.

Piketty, Thomas and Emmanuel Saez. 2006. "The Evolution of Top Incomes: A Historical and International Perspective." *Proceedings of the American Economic Association* 96:200–205.

Raffalovich, Lawrence E. 1993. "Structural Sources of Change in Earnings Inequality: Evidence from the Current Population Survey." *Research in Social Stratification and Mobility* 12:113–44.

Reynolds, Alan. 2007. "Has U.S. Income Inequality Really Increased?" *Policy Analysis* no. 586, January 8. Cato Institute.

Spilerman, Seymour. 1977. "Careers, Labor Market Structure, and Socioeconomic Achievement." *American Journal of Sociology* 85:551–93.

Stolzenberg, Ross M. 1975. "Occupations, Labor Markets and the Process of Wage Attainment." *American Sociological Review* 40:645–65.

Treiman, Donald. 1977. *Occupational Prestige in Comparative Perspective*. New York: Academic Press.

U.S. Census Bureau. 2002. *Design and Methodology*. Current Population Survey. Technical Paper 63RV. Issued March 2002 (http://www.census.gov/prod/2002pubs/tp63rv.pdf).

U.S. Department of Labor. 2006. *Occupation Outlook Handbook, 2006–7 Edition*. Retrieved February 2, 2007 (http://stats.bls.gov/oco/home.htm).

Weeden, Kim A. 2002. "Why Do Some Occupations Pay More than Others? Social Closure and Earnings Inequality in the United States." *American Journal of Sociology* 108:55–101.

Weeden, Kim A., Young-Mi Kim, Mathew Di Carlo, and David B. Grusky. 2007. "Social Class and Earnings Inequality." *American Behavioral Scientist* 50:702–36.

Welch, Finis. 1993. "Matching the Current Population Surveys." *Stata Technical Bulletin*, StataCorp LP, Vol. 2(12).

Western, Bruce and Deirdre Bloome. 2009. "Variance Function Regressions for Studying Inequality." *Sociological Methodology* 39:293–326.

Wooldridge, Jeffrey M. 2002. *Econometric Analysis of Cross Section and Panel Data*. Cambridge, MA: MIT Press.

Wright, Erik Olin and Rachel E. Dwyer. 2003. "The Patterns of Job Expansions in the U.S.A.: A Comparison of the 1960s and 1990s." *SocioEconomic Review* 1:289–325.
Wyatt, Ian D. and Daniel E. Hecker. 2006. "Occupational Changes during the 20th Century." *Monthly Labor Review* (March):35–57.

Ted Mouw is Associate Professor of Sociology at the University of North Carolina at Chapel Hill, and a faculty fellow at the Carolina Population Center. His research interests include social stratification, labor markets, and immigration. He is currently working on models of the intragenerational mobility of low-wage workers (with Arne Kalleberg) and a project on the effect of immigration on the wages of native workers using data from the Longitudinal Employer Household Database (LEHD).

Arne L. Kalleberg is Kenan Distinguished Professor of Sociology at the University of North Carolina at Chapel Hill. He has written extensively on topics related to the sociology of work, organizations, occupations and industries, labor markets and social stratification, and inequality. He is the author of *The Mismatched Worker* (2007). His most recent book examines the rise of polarized jobs and precarious employment relations in the United States since the 1970s. Current projects include studies of precarious work in Asia and in Europe, and (with Ted Mouw) of career mobility and inequality in happiness in the United States.