http://clevelandfed.org/research/review/
Economic Review 1997 Q2

Workforce Composition and Earnings Inequality

by Mark E. Schweitzer

Mark E. Schweitzer is an economist at the Federal Reserve Bank of Cleveland. The author gratefully acknowledges the advice of Trudy Cameron, Ben Craig, Bruce Fallick, David Lewin, Dan Mitchell, and Finis Welch. He also thanks workshop participants at UCLA, Texas A&M, Case Western Reserve University, and the Federal Reserve Banks of Cleveland and Dallas.

Introduction

As the United States passed through another election phase last fall, we again heard about the increasingly unequal earnings prospects of America's workforce. While there are many suggested remedies, making educational opportuniies more available to all is the most common.[1] Proponents of this approach believe that rising returns to education can be attenuated by increasing the supply of highly educated workers and reducing the supply of less skilled workers. This follows from analyses indicating that education is the primary factor contributing to earnings inequality.[2]

Existing research in this area has typically focused on a single demographic group rather than on how demographic groups' earnings relate.[3] Juhn and Murphy (1997) extend their earlier analysis on white males to both sexes by considering the effects of marriage and family structure on family inequality. They find that like workers tend to marry one another, increasing the earnings gap between families. This approach returns the focus to general workforce inequality, but includes four major demographic groups in a generalized inequality decomposition based on estimating their human capital returns independently. The most notable results of this analysis that could not be ascertained in previous research are 1) the increasing share of women in the workforce and their increasing realized tenures have reduced earnings inequality, and 2) a larger portion of the variation in earnings is associated with the changing composition of the workforce, rather than with changing returns to human capital investments.

The three main qualitative results of the existing research are confirmed in Juhn and Murphy's study, although the levels of these factors are altered somewhat by either the different population or the procedures used in their analysis: 1) educational differences are the primary variable associated with rising inequality, 2) industry affiliation and wage differentials are associated with rising inequality, and 3) wage

■ 1 A good summary of these proposals can be found in Freeman (1996) and in the responses to his article.

■ 2 Juhn, Murphy, and Pierce (1993) and Murphy and Welch (1992) make a strong case for focusing on education. For a broader survey of the literature, see Levy and Murnane (1992).

■ 3 Karoly (1992) studies the importance of gender and other factors individually, while Burtless (1990) presents an extensive comparison of wage inequality between men and women.

differences by experience level (age) or region have little impact on inequality. What is new is that Juhn and Murphy's results apply to the entire full-time/full-year workforce, even though that population has changed dramatically (particularly through the addition of more women and minorities and the higher educational levels attained, especially by those two groups). This suggests that typical policy remedies may take a long time to overcome the trend toward increasing inequality.

While the widening disparity in Americans' earnings is a heavily cited and discussed phenomenon, pinpointing its source is a complex exercise. The level of earnings inequality in a society is determined by interactions among many factors. For example, one worker's higher education level may be offset by another's greater experience, yielding no inequality between them. The potential for interactions among earnings factors is large in a diverse workforce, because diversity introduces many potentially offsetting and augmenting sources of inequality. In fact, as previously documented, the wage structure in the United States has been altered along several dimensions over the last two decades: Educational differentials have expanded, experience profiles have steepened, women's wages have drawn closer to men's, and so on.

The potential for interactions among these factors is not merely of academic interest. Such interactions may alter the impact of wage structure changes, including those encouraged by public policy. Again, it is instructive to look at an example. Increasing the educational level of the highest-paid members of a demographic group may boost the earnings disparity within that group while clearly lowering the societal level of inequality. Ultimately, the effect of any change in the earnings structure on earnings inequality depends on the covariation of the altered factors with the other earnings characteristics of the studied population.

To account for interactions among earnings factors, this paper applies a generalized decomposition to Current Population Survey data on the earnings of all full-time/full-year U.S. labor force participants. The decomposition is implied by a model of earnings that encompasses a broad set of variables *simultaneously* in order to describe sources of earnings: education, experience, industry, and region. Furthermore, to better account for the changing composition of the workforce, the model is estimated independently for each of four race/sex groups (minority males, minority females, white males, and white females). This allows distinct wage determination patterns to emerge for these groups, which might alter the covariation of wages between them.

The remainder of the paper is organized as follows. Section I lays out a framework for determining earnings factors in a diverse workforce. Section II decomposes earnings inequality within this framework and extends the framework to consider the role of rising returns versus demographic changes. Section III summarizes and reconsiders policy prescriptions for the increasing earnings gap.[4]

I. Inequality Implications of Earnings Models

The treatment of earnings inequality in this paper follows the approach of Mincer's (1958) seminal work on human capital and the distribution of personal income—a specification that is now typically used in predictive models of earnings. Mincer used his model to stress that inequality due to human capital differences, a fundamental source of earnings inequality, should be separated from other sources of disparity. The result of differences in human capital investment can be summarized by the classic earnings equation, developed in Mincer (1974):

$$(1)\quad \ln W_i = \ln W_{0i} + rS_i + b_1X_i + b_2X_i^2 + v_i,$$

where $\ln W_{0i}$ is the wage for a worker's innate ability, S_i is years of schooling, X_i is years of experience, and v_i includes unobserved individual differences.

Equation (1) is extended below to provide a better fit with the actual experience profile, as suggested by Murphy and Welch (1990). An important extension of Mincer's framework is to allow workers to gain returns for working in their current industry. This is the logical extension of job-specific human capital (Oi [1962]) to industries. The final factor typically included in earnings models (other than race and sex, which receive a more careful treatment below) is the location of an individual's residence.[5] A simple but limiting means of accounting for these effects is to assume that the differences are constant across characteristics. Then, the earnings equation becomes

4 Construction of the data set, which largely follows the "committed worker" restrictions of Juhn, Murphy, and Pierce (1993), is described in the appendix.

5 See Eberts (1989) for a detailed look at regional wage differences.

TABLE 1

Race/Sex-Group Relative Wages and Workforce Shares (percent)

	1972	1978	1984	1990
Level	**Estimated Value of Race/ Sex-Group Differentials**			
White female	–36.88	–33.49	–29.10	–24.69
Minority female	–40.76	–35.84	–32.46	–28.25
Minority male	–20.71	–15.75	–16.87	–13.72
Frequency	**Percentage of Full-Time/ Full-Year Workforce**			
White male	61.20	56.76	53.27	51.15
White female	28.58	31.72	34.24	35.77
Minority female	4.23	5.17	6.10	6.45
Minority male	5.99	6.34	6.39	6.63

NOTE: Percentages are in terms of weekly wages evaluated around the intercept.
SOURCE: Author's calculations.

$$(2) \quad \ln W_i = \ln W_{0i} + b_1 S_i + b_2 X_i + b_3 D_i^{ind} + b_4 D_i^{oth} + v_i,$$

where S_i represents a vector of schooling-level indicators, X_i is a vector of quadratic experience terms, D_i^{ind} represents industry-specific effects, and D_i^{oth} represents regional effects. In the estimation, the rates of return for the earnings factors are allowed to change from year to year. Thus, the value and distribution of these skills and other factors are allowed to vary with shifts in labor supply and demand.

Accounting for Race and Sex Differences

Why account for changes in the racial and sexual composition of the workforce? Since earnings data were first collected, systematic differences in demographic groups' wages have been apparent. Between 1972 and 1990, a large shift occurred in the demographic composition of the full-time/full-year workforce. Table 1 illustrates both of these trends. The estimated differentials represent the coefficients for dummy variables in an earnings equation as specified above, that is, one which controls for experience, education, aggregated industries, and regions. The value of race/sex-group differentials falls by approximately one-third for each of these groups over the 18-year period, while the relative role in the full-time/full-year labor force for all four groups is rising.

Although the specification developed in equation (2) is a standard framework for measuring wage differences, particularly after including race and sex dummy variables, it provides little information about either the sources of race/sex differences or their effects on overall inequality. In particular, if returns to measured skills vary systematically by race/sex groups, then as the composition of the workforce changes, the estimated rates of return would be altered without any variation in the underlying rates of return for specific race/sex groups. A flexible specification that accounts for these differences by allowing complete variation in rates of return for all factors and for the error term by race/sex group is described in equation (3):

$$(3) \quad \ln W_i = \sum_{C \in \{\text{race/sex groups}\}} (\ln W_{0C} + b_{1C} S_i + b_{2C} X_i + b_{3C} D_i^{ind} + b_{4C} D_i^{reg} + v_{iC}),$$

where C indicates the race/sex group of individual i.

Returns to factors could vary by race or sex for several reasons. Returns to observed factors could differ because of qualities unobserved by the econometrician but seen by market participants. Alternatively, race or sex discrimination could be limited to particular sections of the labor market or restricted to certain factors. One clear source of differences in rates of return by race/sex is the variation in actual experience for given levels of potential experience observed in the Current Population Survey. Differing rates of return could also develop as a response to workers' inability to unbundle their set of skills, as shown by Heckman and Scheinkman (1987). They prove that differences in rates of return for observed and unobserved skill factors can vary by group if the proportions of skills vary and workers cannot market their skills separately.

Covariance Structure of the Model

The implications of this model for mean earnings are well known; however, its implications for earnings *inequality* have been applied only infrequently in the recent surge of inequality

TABLE 2

Correlations between Regression Components

1972	Experience	Education	Industry	Region
Experience	1.0 (0.00)			
Education	–0.1738 (0.0001)	1.0 (0.00)		
Industry	0.0721 (0.0001)	–0.0894 (0.0001)	1.0 (0.00)	
Region	0.0020 (0.6947)	0.0405 (0.0001)	0.0527 (0.0001)	1.0 (0.00)
1990	**Experience**	**Education**	**Industry**	**Region**
Experience	1.0 (0.00)			
Education	–0.1595 (0.0001)	1.0 (0.00)		
Industry	0.0928 (0.0001)	–0.1204 (0.0001)	1.0 (0.00)	
Region	0.0051 (0.2574)	0.0360 (0.0001)	0.0234 (0.0001)	1.0 (0.00)

NOTE: Figures in parentheses are probability values for the null hypothesis that the correlations are zero.
SOURCE: Author's calculations.

literature.[6] Consider a scenario of increasing returns to a single factor—education, for example. The standard decomposition of inequality by subgroups provides a simple comparison of mean earnings by industry and concludes that inequality rises. In terms of equation (2), one treatment of this hypothesis is that the range of vector b_1 is increased, as measured by $\max(b_1) - \min(b_1)$. This raises the variance of the term b_1S_i, but the effect of increasing the range of b_1 on the variance of earnings also depends on the signs of the covariances.

$$(4)\quad \frac{\partial \operatorname{var}(\ln W_i)}{\partial \operatorname{range}(b_1)} = \frac{\partial \operatorname{var}(b_1S_i)}{\partial \operatorname{range}(b_1)} + 2\,\frac{\partial \operatorname{cov}(b_1S_i\,,\, b_2X_i)}{\partial \operatorname{range}(b_1)} + \ldots + 2\,\frac{\partial \operatorname{cov}(b_1S_i\,,\, v_i)}{\partial \operatorname{range}(b_1)}\,.$$

Only the first and last terms of this derivative may be signed: The first is unambiguously positive, and the last (the covariance with the error term) is always zero by ordinary least squares.

Empirically, these covariances are a substantial and statistically significant portion of total wage variation, as indicated by the correlations in table 2. The reported correlations are for a regression of individual log wages on four of the variable categories discussed throughout this paper: experience, education, industry, and region. In addition to being generally significant, these correlations may change over time, as a cursory comparison of the 1972 and 1990 results indicates. Individual returns to education appear to be especially correlated with two other recognized earnings factors: experience and industry. This is not surprising, since educational levels are higher for younger cohorts, and education is clearly associated with one's industry choice.

Neglecting the covariances among explanatory variables affects the interpretation of the impact of industry wage differentials on earnings inequality. For example, Freeman (1991) argues that the loss of labor union premiums for low-skilled workers has exacerbated U.S. earnings inequality. Standard subgroup decompositions would be inappropriate without including other observed determinants of industry wage differentials, since they would indicate only the effect of union wage differentials. Freeman's point is that inequality is lower because of a negative covariance between union effects and skill factors.[7] An inequality decomposition should account for this negative covariance, thereby reducing the earnings inequality associated with union wage differentials. Without direct observation of union status, union effects can be viewed as a component of $b_3D_i^{ind}$, and the argument applies to industry premiums as well. Accounting for covariances can be similarly justified for most factors considered in the earnings inequality literature.

Equation (3) shows the more complicated covariance structure to be summarized by the decomposition. This extension alters the interpretation of the factors and allows for comparisons across groups. A change in the rate of return a single group is paid for a factor depends on both the covariance structure with that group's other factors and the covariances between that group's and other groups'

6 Smith and Welch (1979) did recognize the importance of covariances between explanatory variables in their analysis of race differences in earnings inequality. A similar technique is applied to prime-age white males in Blackburn (1990).

7 Freeman avoids this criticism by not performing an explicit inequality decomposition. Instead, he applies shift/share analysis to regression estimates after controlling for education.

TABLE 3

Education Differentials by Race/Sex Group (percent)

	Estimated Value of Education Differentials			
1972	**White Men**	**Minority Men**	**White Women**	**Minority Women**
High school dropout	–18.22	–15.99	–11.19	–13.28
Some college	14.63	7.29	12.48	9.15
College graduate	50.54	44.52	51.71	72.20
Post-graduate	75.49	77.01	80.16	124.49
1990	**White Men**	**Minority Men**	**White Women**	**Minority Women**
High school dropout	–25.72	–15.42	–21.42	–15.06
Some college	19.01	20.12	18.26	15.47
College graduate	58.69	51.69	65.02	52.29
Post-graduate	89.47	99.58	102.63	102.20

NOTE: Percentages are in terms of weekly wages evaluated around the race/sex-group intercept.
SOURCE: Author's calculations.

FIGURE 1

Experience–Earnings Profiles: 1972

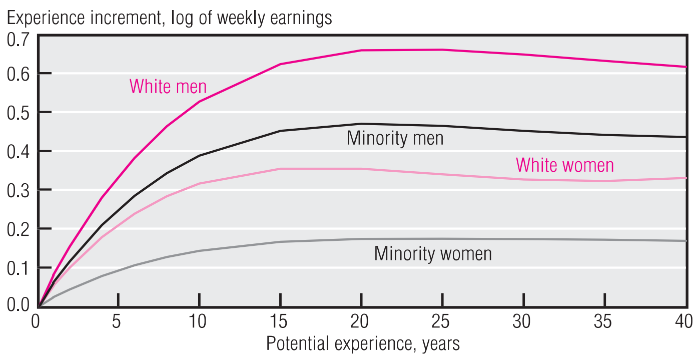
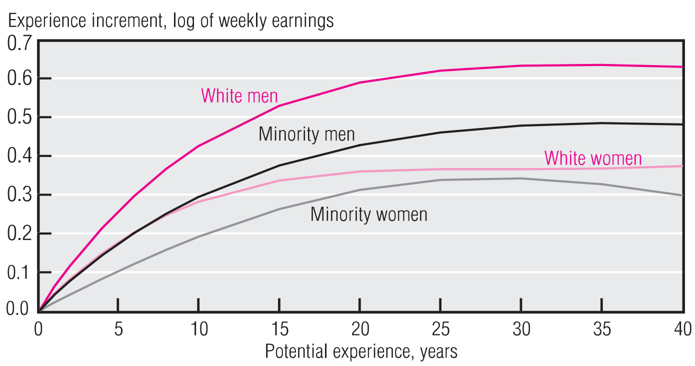


FIGURE 2

Experience–Earnings Profiles: 1990



SOURCE: Author's calculations.

wages. Repeating the earlier example of an increase in the rate of return to education for workers in group 1 of four race/sex groups, we have

$$(5)\quad \frac{\partial\, \mathrm{var}(\ln W_i)}{\partial \mathrm{range}(b_{1,1})} = \frac{\partial\, \mathrm{var}(b_{1,1}S_i)}{\partial \mathrm{range}(b_{1,1})} + 2\,\frac{\partial\, \mathrm{cov}(b_{1,1}S_i\,,\, b_{2,1}X_i)}{\partial \mathrm{range}(b_{1,1})} + \ldots + 2\,\frac{\partial\, \mathrm{cov}(b_{1,1}S_i\,,\, v_i)}{\partial \mathrm{range}(b_{1,1})} + 2 \sum_{j=2,3,4} \frac{\mathrm{cov}(b_{1,1}S_i\,, \ln W_{i\in j})\,\cdot}{\mathrm{range}(b_{1,1})}$$

Equation (5) raises the possibility that certain factors could increase earnings inequality within a group and yet reduce populationwide inequality.

Race/sex groups have different returns to factors, and these returns change across the period. Educational differentials are shown in table 3. With the exception of highly educated minority women, additional education is better rewarded in 1990 than in 1972 for all race/sex groups. Interestingly, education was more steeply rewarded among women than men in both years, indicating that in terms of educational differentials, men's wages have shifted toward the steeper profile of women's wages. The impact of these increasing returns to education on overall wage inequality could be mixed, since growth in higher-education differentials for women could reduce earnings inequality because of their generally lower wages.

The other explanatory variable that is estimated quite differently for each race/sex group is returns to potential experience. "Potential" is stressed here because actual experience levels associated with years of potential experience may vary sharply, particularly for women in the early years of the period. One difficulty with experience returns is pinpointing the size and location of the important differences that contribute to earnings inequality. By plotting the full experience–earnings profiles for each of the groups, the scale of the differences can be evaluated at any level of potential experience. Figures 1 and 2 describe the rates of return to experience at the characteristic means of the race/sex groups for 1972 and 1990. Notice that both figures indicate substantial differences; however, the plots converge noticeably over the period.

These two explanatory variables are the most obvious reasons to embark on a decomposition that can account for differing wage structures by race/sex groups, but other differences also exist (for example, demographic differences in the industry-specific terms).

The Decomposition

This paper applies an alternative inequality decomposition that utilizes our understanding of the sources of earnings differences. It uses estimates from standard semilog earnings models to separate earnings into additive components, which can then be evaluated as separate earnings factors. This approach offers several advantages: First, the decomposition can be based on models that have long been accepted by labor economists as reasonably accurate representations of individual earnings. Second, inequality can be speedily decomposed into many categories. Third, inequality can be decomposed according to both discrete and continuous variables.

The decomposition is specified by the following, where Y_i is actually the sum of k component incomes measured in logs, and Y_i^k identifies the k^{th} income component:

$$(6)\quad \sigma^2(Y) = \sum_{i=1}^{N} \frac{(Y_i - \mu)}{n} \sum_{k=1}^{K} Y_i^k$$

$$= \sum_{k=1}^{K} \mathrm{cov}\,(Y_k, Y) \equiv \sum_{k=1}^{K} S_k^*(\sigma^2).$$

The term $S_k^*(\sigma^2)$ follows Shorrocks' (1982) notation for the k^{th} earnings-component-decomposition term of the variance (σ^2), which is measured by the covariance of the income components and total incomes. Shorrocks develops the variance for expository purposes only and does not discuss the variance of log earnings (LV), since it does not satisfy the principle of transfers—a criticism that Creedy (1977) has shown to be irrelevant within the ranges of income or earnings seen in developed economies.[8]

For the simplest case where the earnings factors can be described as partitioned and complete sets of dummy variables, the decomposition on those factors is equivalent to between-group components of a subgroup decomposition on those subgroups. Consider a population that can be divided into N subgroups. Standard coding of the dummy variables results in an X matrix of

$$X = \begin{bmatrix} i_1 & 0 & 0 & \cdots & 0 \\ i_2 & i_2 & 0 & & 0 \\ i_3 & 0 & i_3 & & 0 \\ \vdots & & & & \vdots \\ i_N & 0 & \cdots & & i_N \end{bmatrix},$$

where i_j represents vectors of ones of length n_j, which is the number of members in group j.[9] This matrix excludes the first group from the dummy variables to avoid linear dependence with the intercept. Regression of a vector $Y = (y_i)$ on X results in the following coefficients of β and predictions of $X\beta$:

$$\beta = \begin{bmatrix} \bar{y}_1 \\ \bar{y}_2 - \bar{y}_1 \\ \vdots \\ \bar{y}_n - \bar{y}_1 \end{bmatrix} \quad \text{and} \quad X\beta = \begin{bmatrix} \bar{y}_1 \\ \bar{y}_2 \\ \vdots \\ \bar{y}_n \end{bmatrix},$$

where $\bar{y}_j$ is a vector of the j^{th} group mean. Note that this regression is just another way to calculate the group means.

Treating $X\beta$ and $Y - X\beta$ as factors of the total (Y) and applying the formula for factor decomposition of the variance (equation [6]) result in a standard variance decomposition by subgroups:

$$(7)\quad \mathrm{cov}(X\beta, Y) = \sum_{i=1}^{N} \left[\frac{1}{N}\, \bar{y}_{j(i)}\, y_i - E(\bar{y}_{j(i)})\bar{y} \right]$$

$$= \sum_{j=1}^{J} \left[\frac{1}{N}\, \bar{y}_j \sum_{i \in j} y_i \right] - \bar{y}^2$$

$$= \sum_{j=1}^{J} \frac{n_j}{N} (\bar{y}_j^2 - \bar{y}^2).$$

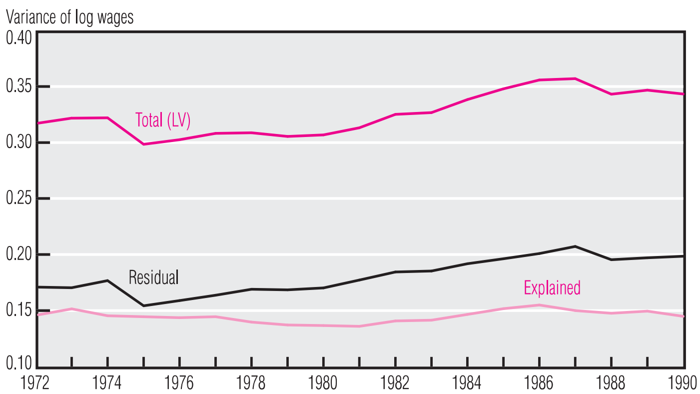
If Y is wages measured in logs, then the final term of equation (7) is the sum of j between-group terms of the subgroup decomposition of the LV. The within-group portion is simply var $(Y) - \mathrm{cov}(X\beta, Y)$.

8 The principle of transfers requires the inequality measure to increase whenever income is transferred from a poorer to a richer person.

9 The estimations reported later in this paper are based on regressions in which the dummy variables are effects coded. With effects coding, the coefficients express the difference between group k's mean and all groups' mean wages. The results reported here hold for both standard dummy-variable coding and effects coding, as long as the dummy variables serve as a complete set of group-specific intercepts.

FIGURE 3

Total, Explained, and Residual Earnings Inequality



SOURCE: Author's calculations.

The weighting scheme is identical in the simplest case, but it is easy to move the component decomposition toward simultaneous estimation of a variety of factors affecting earnings. Accounting for covariances can be described as controlling for other effects in a regression framework. The value of this can be seen by considering an example. Wages in many service industries have remained steady or have grown relative to manufacturing wages.[10] This could be due to service industries paying a greater industry differential or to their hiring more-skilled workers (in particular, more-educated workers). In either case, controlling for the level of education and experience in the service workforce would identify lower relative service industry wages. A traditional subgroup decomposition on industries would miss the shift in wage differentials that is hidden by skill upgrading in this sector.

II. Inequality Decompositions

Figure 3 shows the degree to which the model is able to predict observed inequality differences. This is the simplest decomposition possible, but it provides information on how completely the model represents the data. If this were a single-equation model, the percentage of predicted inequality explained would be equal to the R^2 of a regression. Thus, earnings models should not be expected to describe all (or even most) of the variation in earnings when a plethora of important but unobserved individual differences is not taken into account. For 1975, the model predicted an inequality level of 0.1382, which is 46.29 percent of all variation in log earnings. The model predictions of the economywide LV level are relatively stable at around 0.14. However, inequality due to the residual widens throughout the period; thus, the model explains a declining share of the LV of wages.

In addition, the shift in imputation techniques used by the Census Bureau appears to be concentrated in the residuals, which fall from their trend in 1975. At this level of decomposition, the trend in the observed portion of earnings inequality is maintained through the switch in techniques, while the residual portion is dramatically altered.

Factor Shares of Explained Earnings Inequality

This earnings-component-based method of decomposition can be easily applied to any collection of the model's set of variables. Although the overall model's explained inequality changed little from 1972 to 1990, the effects of certain worker characteristics rose or fell rapidly. The results of decomposing the model's estimates into categories are shown in table 4. The experience group includes the quartic terms of potential experience. The education group includes the dummy variables for high school dropout, some college, college graduate, and postgraduate. The race/sex group is implied by the constants of the race/sex-group earnings equations, which are the baseline earnings of individuals of that group after controls have been applied. The industry group includes the 38 industry dummy variables. The region group includes dummy variables for the nine U.S. census regions. The estimated wage effects ($X\beta$) are calculated for each group of variables from annual regressions.

The experience group is a key factor in the explained variation early in the period, reaching 0.0715 LV (or 49.2 percent of explained inequality) in 1974, but its influence declines

10 Average hourly earnings for manufacturing workers fell from \$8.33 in 1970 to \$8.07 in 1990 (1982 dollars). Over the same period, service industry wages rose from \$6.99 to \$7.39 (see 1992 *Statistical Abstract of the United States,* table 650, p. 410).

TABLE 4

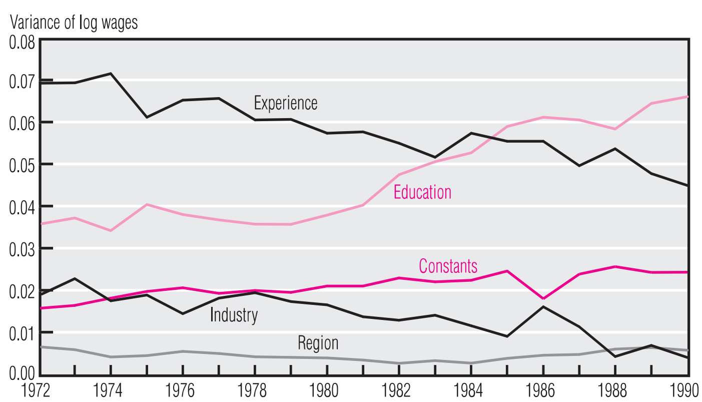
Estimated Earnings Components with Independent Race/Sex Groups

	Levels of Explained Inequality When Factor Returns Differ by Race/Sex Groups					
Year	Total Explained	Experience	Education	Race/ Sex	Industry	Region
1972	0.1463	0.0692	0.0358	0.0191	0.0157	0.0065
1973	0.1514	0.0693	0.0372	0.0227	0.0164	0.0058
1974	0.1454	0.0715	0.0342	0.0175	0.0181	0.0041
1975	0.1444	0.0611	0.0403	0.0188	0.0197	0.0044
1976	0.1436	0.0652	0.0380	0.0144	0.0206	0.0054
1977	0.1446	0.0656	0.0367	0.0181	0.0192	0.0049
1978	0.1397	0.0605	0.0357	0.0194	0.0199	0.0042
1979	0.1371	0.0606	0.0357	0.0173	0.0195	0.0040
1980	0.1366	0.0573	0.0379	0.0165	0.0210	0.0039
1981	0.1359	0.0576	0.0403	0.0137	0.0210	0.0034
1982	0.1408	0.0550	0.0475	0.0129	0.0229	0.0026
1983	0.1415	0.0516	0.0506	0.0140	0.0220	0.0032
1984	0.1465	0.0573	0.0527	0.0115	0.0224	0.0027
1985	0.1517	0.0554	0.0589	0.0090	0.0245	0.0038
1986	0.1550	0.0554	0.0611	0.0160	0.0180	0.0045
1987	0.1499	0.0496	0.0605	0.0113	0.0238	0.0047
1988	0.1477	0.0536	0.0583	0.0042	0.0256	0.0060
1989	0.1496	0.0477	0.0644	0.0068	0.0242	0.0064
1990	0.1449	0.0449	0.0660	0.0040	0.0243	0.0057

SOURCE: Author's calculations.

FIGURE 4

Estimated Earnings Components with Independent Race/Sex Groups



SOURCE: Author's calculations.

thereafter, bottoming out at 0.0449 LV (31.0 percent) in 1990. This is confirmed in figure 4, which compares the relative trends of all five factors. Recalling the experience returns shown in figures 1 and 2, we see that these differences likely reflect the race/sex composition of the labor force as much as they do cohort differences. Previous work with a single-equation model revealed little trend in the experience factor. As the returns to experience converge, but steepen, across demographic groups, the contribution of the earnings factor to inequality declines. This also suggests that it is not experience-profile differences across ages that drive this result, but differences in potential experience returns across groups, because the returns become steeper for all groups by 1990. This is clearly a case where a factor that on its own would contribute to rising inequality (a steeper experience profile) reduces inequality across groups.

Education variables explain a much larger share of the variance of log earnings in recent years. The explained variance accounted for by education dummies rises from a low of 0.0342 LV (23.5 percent of explained inequality) in 1974 to a high of 0.0660 LV (45.6 percent) in 1990. The explanatory power of the education variables increases sharply from the mid-1970s on. Unlike the results for the experience terms, however, rising differentials for all groups (shown in table 3) add to inequality, rather than offsetting other differences. This makes educational levels stand out as a source of inequality that spans different demographic groups in a way that does not ameliorate inequality levels associated with other factors, including the unobserved factors in the residual. It should be noted that while the differentials are certainly important, the fraction of the workforce attaining higher educational levels has also risen.

The race/sex term is defined by the constants of the regression equations. It thus represents baseline differences not associated with return differences on the included factors, rather than an inclusive measure of group differences. It starts with a peak explanatory power of 0.0227 (15.0 percent of explained earnings inequality) in 1973, but by 1990 accounts for only 0.0040 LV (2.7 percent). This dramatic decline, which is spread over the period, has not been noted in previous studies because most researchers either have considered only men or have treated men and women as if they participated in different labor markets. Combined with the reduced effect of experience as a factor in inequality, factors closely related to differences among demographic groups have declined considerably as sources of inequality.

The share of industry variables in explained inequality is not as large or as steeply trended as either the education or race/sex shares. However, the effect of industry wage differentials rises from 0.0157 LV (10.7 percent of explained earnings inequality) in 1972 to 0.0256 LV (17.3 percent) in 1988. The share of inequality represented by the industry factor does little to bolster theories positing that the increase in overall inequality results mainly from industrial shifts. However, unlike in previous studies, the trend in the industry component is noticeably upward and economically significant. Regional differences play a consistently small role in earnings inequality, reaching a low of 0.0026 LV (1.9 percent) in 1982.

Simple calculations from table 4 indicate that trends in some of these factors are quite large. Inequality due to educational differences rose 116.2 percent more than overall earnings inequality among full-time/full-year labor force participants from 1972 to 1990. This implies that if all factors other than education (including the residuals) were held constant over the period, earnings inequality would have risen 16.2 percent more than it actually did. Even the relatively small industry factor grew 33.1 percent as much as overall inequality.

These increases are more than offset by the drop in the experience and race/sex factors. Experience-related inequality declined at a rate equal to 93.5 percent of the increase in overall earnings inequality over the 1972–90 period, and the race/sex factor fell 58.1 percent versus the same measure. These factors, combined with the small regional factor, yield explained inequality levels that actually decrease as overall inequality and inequality associated with education and industry affiliations rise.

Fixed-Return Comparisons

A valuable extension of the preceding analysis is to separate the effects of population shifts from the effects of changes in returns to worker characteristics. Basic shift/share analysis, in which a population having given characteristics is adjusted in order to isolate the population effects, cannot be applied to this decomposition because the correlations of individual characteristics at the observation level are critical. Shift/share analysis implicitly assumes that the nature of the correlations stays constant.

A related approach is to contrast the explained inequality level under the restriction that the estimated coefficients are constant in all years. Here, the restricted case is referred to as fixed-return estimates because the coefficients represent the amount a hypothetical average individual is paid for having that characteristic. This comparison can isolate the effects of changes in rates of return paid to earnings factors from the changing distribution of those factors. Much as in shift/share analysis, in addition to the returns and quantity terms, there is a covariance between the two terms that is assumed to be zero. This allows for the simple separation

$$(8) \qquad S_k^*(\sigma^2) = \text{cov}(X_k \beta_k, Y_i) = \text{cov}\,(X_k \tilde{\beta}_k, Y_i) + \text{cov}\,[X_k(\beta_k - \tilde{\beta}_k),\, Y_i],$$

where $\tilde{\beta}_k$ represents any desired value of the coefficient vector for the k^{th} factor.

Table 5 shows the difference between the restricted (coefficients maintained at 1972 levels) and unrestricted inequality components over time. The difference between the two estimates equals the final term in equation (8), which is an inequality-weighted measure of the difference between coefficients. A positive value indicates that allowing the coefficients to vary increases inequality; a negative number implies reduced disparity when coefficients are allowed to change.

If all returns to worker characteristics were held at their 1972 levels, the explained level of earnings inequality would have been slightly lower in 1990 than in 1973. This suggests that overall shifts in the composition, without any change in the earnings functions, has raised earnings inequality. Referring again to table 5, with constrained rates of return, earnings inequality would have been higher, with 1972's return levels, from 1974 to 1983. The reversal of this result is due to sharply rising returns to education in the 1980s. The largest differences, and therefore the largest return-related shifts, occurred in the education-, race/sex-, and industry-related components.

The change in experience-related inequality is 73 percent larger when determined without any change in relative returns, whereas race/sex-related inequality drops off only 49 percent as much. Education-related inequality is even more affected by shifts in the number of workers at various schooling levels than by shifts in the returns for increases in that component (64 percent of the change would have occurred with no change in returns). By contrast,

TABLE 5

Effect of Holding Returns Constant with Independent Race/Sex Groups

	Increase in Factor Estimates with Flexible Returns					
Year	Total Explained	Experience	Education	Race/ Sex	Industry	Region
1972	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
1973	0.0029	–0.0025	0.0007	0.0033	0.0018	–0.0005
1975	–0.0003	0.0004	–0.0014	–0.0013	0.0034	–0.0014
1976	–0.0009	–0.0064	0.0023	0.0009	0.0036	–0.0013
1977	–0.0009	–0.0029	0.0012	–0.0031	0.0041	–0.0002
1978	–0.0008	–0.0032	–0.0001	0.0005	0.0023	–0.0002
1979	–0.0024	–0.0065	–0.0007	0.0019	0.0034	–0.0005
1980	–0.0029	–0.0050	–0.0010	0.0000	0.0036	–0.0005
1981	–0.0035	–0.0073	–0.0001	–0.0002	0.0047	–0.0008
1982	–0.0041	–0.0060	0.0008	–0.0027	0.0047	–0.0010
1983	–0.0008	–0.0069	0.0041	–0.0029	0.0064	–0.0015
1984	0.0002	–0.0086	0.0055	–0.0010	0.0056	–0.0013
1985	0.0027	–0.0038	0.0056	–0.0034	0.0061	–0.0018
1986	0.0062	–0.0046	0.0099	–0.0055	0.0076	–0.0011
1987	0.0070	–0.0044	0.0099	0.0019	0.0005	–0.0009
1988	0.0044	–0.0080	0.0088	–0.0028	0.0069	–0.0006
1989	0.0034	–0.0022	0.0071	–0.0091	0.0075	0.0001
1990	0.0050	–0.0063	0.0100	–0.0060	0.0067	0.0007

SOURCE: Author's calculations.

virtually all of the rise in the industry component is driven by shifts in industry wage differentials, rather than by shifts in industry employment shares.

The fact that much of the change in earnings inequality occurs without changes in relative earnings is significant. A large part of the change in demographics is predictable. We know the basic characteristics of people poised to enter the labor force, and we can guess that trends in industry employment shares are likely to continue for several years. The retiring labor force in the United States is more male, more white, less educated, and more likely to work in manufacturing industries. Replacing these workers implies a continuation of offsetting compositional changes on the earnings inequality of the workforce that, depending on which effect dominates, will determine much of the inequality trend into the next decade.

III. Conclusion

Stepping back from the technical details of the decompositions, we can see the complexity involved in addressing total earnings inequality as a public policy issue. While policies might be easily structured to benefit particular groups of workers, important covariances in earnings factors across groups can lead to changes in overall inequality that are either positive or negative. The decomposition employed in this paper can be used to verify or alter findings based on studies of single demographic groups.

Notably, the growing importance of educational differences is verified across all four demographic groups examined here, despite their widely varying schooling levels. The rise in education-related inequality, which is generally ascribed to rising returns, appears to be more than 50 percent determined by the size of the highly educated labor force, at least in this sample. Neglecting the participation of a growing fraction of the labor force may have caused previous researchers to focus excessively on shifts in returns.

The analysis also establishes the direct role of changing workforce demographics. Race/sex differentials have contributed far less to recent inequality levels than was historically the case, masking part of the widening disparity in other factors. These trends are driven both by changes in relative pay rates and shifts in the composition of the labor force. The largest factor is differing rates of return on potential experience by race/sex group. Declines in this factor have resulted primarily from changing participation rates, not from shifts in the observed experience–earnings profiles.

The decompositions presented here generally point to a larger role for the composition of the full-time/full-year workforce than has previous research. Policy prescriptions based on the existing literature tend to ignore the effects of this striking change. While such remedies may still be appropriate, the fact that much of the inequality trend has been driven by changes in the composition of the U.S. labor force suggests that policies which alter the returns to schooling or other human capital factors will take a long time to work. One reason is that the composition changes realized over the last decade are likely to continue, if only because entering generations are replacing retiring workers who possess characteristics much more typical of the earliest periods of this analysis.

Appendix: The Data Set

The data set is derived from the March Current Population Surveys (CPS) spanning the years 1973 to 1991. Every month, the U.S. Census Bureau interviews about 58,000 households (including approximately 122,000 persons age 14 and over) as part of the CPS. Each sample is designed to be representative of the civilian, noninstitutional population. The March surveys throughout this period include information on individuals' personal characteristics (age, sex, race, and education) and on their residence and employment during the previous year (total wages and salaries, weeks worked, hours worked per week, industry, and occupation). These features have made the March supplement the primary data source used in earnings distribution analyses.

I selected individuals who showed strong attachment to the labor force. The sample includes civilians over age 16 who are not self-employed and who missed no weeks of work because of schooling or retirement. It is further limited to workers who reported being in the labor force (working or unemployed) at least 39 weeks and who worked full time (at least 35 hours per week) in the previous year. Although designed to be similar to the sample used by Juhn, Murphy, and Pierce (1993), mine includes both male and female workers of all races in order to paint a more complete picture of the labor market.

Certain adjustments to the earnings data were also necessary. Top-coded data were assigned the truncated mean earnings implied by a Pareto distribution based on the highest reported earnings. Observations with real weekly wages of less than half the 1982 minimum wage for a full-time job were dropped because they are likely to be faulty. Juhn, Murphy, and Pierce show that differences in the imputation techniques used by the Census Bureau can alter wage inequality, but that these differences are largely limited to extremes of the distribution. The biggest switch occurs between 1974 and 1975 and is visible in the decompositions reported here. To isolate the conclusions of this paper from the issues that affect the fringes of the distribution, the analysis was also completed with a sample from which the top and bottom 5 percent of earners were removed.[11] There were no differences in the truncated sample analysis that would alter the conclusions of this paper.

The Census Bureau changed its industry codes twice during the sample period. However, the basic structure of the industry coding system was not altered at the two-digit level and could be mapped into consistent two-digit Standard Industry Codes. I aggregated some of these codes in order to reduce the number of industries to a manageable number (39) and to increase the cell sizes for small industries.

11 Neither top-coded data nor subminimum wage earnings were ever more than 5 percent of my sample.

References

Blackburn, M.L. "What Can Explain the Increase in Earnings Inequality among Males?" *Industrial Relations*, vol. 29, no. 3 (Fall 1990), pp. 441–56.

Burtless, G. "Earnings Inequality over the Business and Demographic Cycles," in G. Burtless, ed., *A Future of Lousy Jobs?* Washington, D.C.: The Brookings Institution, 1990, pp. 77–117.

Creedy, J. "The Principle of Transfers and the Variance of Logarithms," *Oxford Bulletin of Economics and Statistics*, vol. 39, no. 2 (May 1977), pp. 153–58.

Eberts, R.W. "Accounting for the Recent Divergence in Regional Wage Differentials," Federal Reserve Bank of Cleveland, *Economic Review*, vol. 25, no. 3 (Quarter 3 1989), pp. 14–26.

Freeman, R.B. "How Much Has De-Unionization Contributed to the Rise in Male Earnings Inequality?" National Bureau of Economic Research, Working Paper No. 3826, August 1991.

________. "Toward an Apartheid Economy?" *Harvard Business Review*, September/October 1996, pp. 114–21.

Heckman, J., and J. Scheinkman. "The Importance of Bundling in a Gorman–Lancaster Model of Earnings," *Review of Economic Studies*, vol. 54, no. 2 (April 1987), pp. 243–55.

Juhn, C., and K.M. Murphy. "Wage Inequality and Family Labor Supply," *Journal of Labor Economics*, vol. 15, no. 1, part 1 (January 1997), pp. 72–97.

__________, __________, and B. Pierce. "Wage Inequality and the Rise in Returns to Skill," *Journal of Political Economy*, vol. 101, no. 3 (June 1993), pp. 410–42.

Karoly, L.A. "Changes in the Distribution of Individual Earnings in the United States: 1967–1986," *Review of Economics and Statistics*, vol. 74, no. 1 (February 1992), pp. 107–15.

Levy, F., and R.J. Murnane. "U.S. Earnings Levels and Earnings Inequality: A Review of Recent Trends and Proposed Explanations," *Journal of Economic Literature*, vol. 30, no. 3 (September 1992), pp. 1333–81.

Mincer, J. "Investment in Human Capital and Personal Income Distribution," *Journal of Political Economy*, vol. 66, no. 4 (August 1958), pp. 281–302.

________. *Schooling, Experience, and Earnings.* New York: National Bureau of Economic Research, 1974.

Murphy, K.M., and F. Welch. "Empirical Age–Earnings Profiles," *Journal of Labor Economics*, vol. 8, no. 2 (April 1990), pp. 202–29.

________, and ________. "The Structure of Wages," *Quarterly Journal of Economics*, vol. 107, no. 1 (February 1992), pp. 285–326.

Oi, W.Y. "Labor as a Quasi-Fixed Factor," *Journal of Political Economy*, vol. 70, no. 6 (1962), pp. 538–55.

Shorrocks, A.F. "Inequality Decomposition by Factor Components," *Econometrica*, vol. 50, no. 1 (January 1982), pp. 193–211.

_______. "Inequality Decomposition by Population Subgroups," *Econometrica*, vol. 52, no. 6 (November 1984), pp. 1369–85.

Smith, J., and F. Welch. "Inequality: Race Differences in the Distribution of Earnings," *International Economic Review*, vol. 2, no. 2 (June 1979), pp. 515–26.