

**ACCOUNTING FOR INCOME INEQUALITY AND ITS  
CHANGE: A NEW METHOD, WITH APPLICATION TO  
THE DISTRIBUTION OF EARNINGS IN  
THE UNITED STATES**

**Abbreviated Title: Accounting for Income Inequality**

by

**Gary S. Fields**  
**Professor, Cornell University**  
**School of Industrial Relations**  
**Ithaca, NY 14853-3901 U.S.A.**  
**Telephone: 607-255-4561**  
**Fax: 607-255-4496**  
**[gsf2@cornell.edu](mailto:gsf2@cornell.edu)**

**October, 2002**

**Forthcoming: Research in Labor Economics**

## **ABSTRACT**

This paper devises a new method for using the information contained in income-generating equations to "account for" or "decompose" the level of income inequality in a country and its change over time. In the levels decomposition, the shares attributed to each explanatory factor are independent of the particular inequality measure used. In the change decomposition, methods are presented to break down the contribution of each explanatory factor into a coefficients effect, a correlation effect, and a standard deviation effect. In an application to rising earnings inequality in the United States, it is found that schooling is the single most explanatory variable, only one other variable (occupation) has any appreciable role to play, and all of schooling's effect was a coefficients effect.

**ACCOUNTING FOR INCOME INEQUALITY AND ITS CHANGE:  
A NEW METHOD, WITH APPLICATION TO  
THE DISTRIBUTION OF EARNINGS IN  
THE UNITED STATES**

**1. Introduction**

For decades, economists and other social scientists have sought to understand the inequality of income (or earnings or wages) using regression models.<sup>1</sup> Typically, the logarithm of the income of individual  $i$  in country/group/time  $t$  is regressed on a number of explanatory variables. Assuming that these have been chosen carefully in light of theory and past empirical findings, the question then is how to use the information contained in such income-generating equations to "account for" or "decompose" income inequality.<sup>2,3</sup>

This paper proposes a new methodology for answering two questions.<sup>4</sup> First, given an income-generating function estimated by a standard semi-log regression, how much income inequality is accounted for by each explanatory factor? This shall be termed the "levels question," the answer to which is of the form "x% of the inequality of income is attributable to education, y% to region, z% to gender, etc." Second, denoting the two countries, groups, or dates by 1 and 2 respectively, given estimates of comparable income-generating functions

$$\ln(Y_{i1}) = \alpha_1 + \sum_j \beta_{j1} x_{ij1} + \varepsilon_{i1} \quad (1)$$

and

$$\ln(Y_{i2}) = \alpha_2 + \sum_j \beta_{j2} x_{ij2} + \varepsilon_{i2}, \quad (2)$$

how much of the difference in income inequality between one country and another, between one group and another within a country, or between one date and another is accounted for by education, by potential experience, and by the other explanatory factors? This shall be called the "differences question."<sup>5</sup>

Past literature, reviewed in Sections 2.D and 3.D, provides approximate regression-based answers to these two questions, but until now, no exact decomposition has been available. This paper shows that such a decomposition can be gotten and further that under a quite acceptable set of assumptions, the percentage contribution of a given explanatory factor  $x_j$  at time  $t$  is independent of which inequality measure is chosen. The new procedure for the levels question is presented in Section 2 and for the difference question in Section 3.

Section 4 applies these methodologies to quantify the role of different explanatory factors in accounting for levels of earnings inequality at a point in time and the change in earnings inequality over the last twenty years in the United States. Section 5 sums up.

Before proceeding, it should be noted that the methods developed in this paper are quite general. Although the motivating questions and the methods derived are presented in terms of inequality of income, the same techniques are applicable to answering the levels question and the difference question for any continuous variable.

## **2. Accounting for Income Inequality at a Point in Time**

In this section, a method is proposed to account the inequality of income in a single survey. Start with an income-generating function, based on human capital theory or some other underlying theoretical model, in which income is a function of a certain number of "variables" or "factors." The decomposition is based on the income-generating function (1), which can be rewritten as

$$\ln Y_{it} = a_t' Z_{it} \quad (3.a)$$

where

$$a_t = [\alpha_t \ \beta_{1t} \ \beta_{2t} \ \dots \ \beta_{Jt} \ 1] \quad (3.b)$$

and

$$Z_{it} = [1 \ x_{i1t} \ x_{i2t} \ \dots \ x_{iJt} \ \varepsilon_{it}] . \quad (3.c)$$

On the assumption that "good" estimates have been gotten for the coefficients on the variables, the strategy for deriving a useful decomposition equation is first to decompose the log-variance of income and then to show that the same decomposition applies to other inequality measures as well. <sup>6</sup>

### **A. Decomposing the Log-variance**

Starting with the income-generating functions (3.a-c), take the variance of both sides. On the left-hand side is a simple measure of inequality, the log-variance. The variance of the right-hand side can be manipulated using the following:

**Theorem (Mood, Graybill, and Boes):** Let  $A_1, \dots, A_P$  and  $B_1, \dots, B_Q$

be two sets of random variables, and let  $a_1, \dots, a_P$  and  $b_1, \dots, b_Q$

be two sets of constants. Then

$$\text{cov} \left[ \sum_{p=1}^P a_p A_p, \sum_{q=1}^Q b_q B_q \right] = \sum_{p=1}^P \sum_{q=1}^Q a_p b_q \text{cov}[A_p, B_q] . \quad (4)$$

Applying this theorem in the context of a single random variable  $\ln Y$  such that

$$\ln Y = \sum_{j=1}^{J+2} a_j Z_j ,$$

we have

$$\text{cov} \left[ \sum_{j=1}^{J+2} a_j Z_j, \ln Y \right] = \sum_{j=1}^{J+2} \text{cov}[a_j Z_j, \ln Y] . \quad (5)$$

But because the left-hand side of (5) is the covariance between  $\ln Y$  and itself, it is simply the variance of  $\ln Y$ . Thus,

$$\sigma^2(\ln Y) = \sum_{j=1}^{J+2} \text{cov}[a_j Z_j, \ln Y] , \quad (6.a)$$

or, upon dividing through by  $\sigma^2(\ln Y)$ ,

$$100\% = \frac{\sum_{j=1}^{J+2} \text{cov}[a_j Z_j, \ln Y]}{\sigma^2(\ln Y)} \equiv \sum_{j=1}^{J+2} s_j(\ln Y), \quad (6.b)$$

where each  $s_j(\ln Y)$  is a so-called "relative factor inequality weight" given by

$$s_j(\ln Y) = \text{cov}[a_j Z_j, \ln Y] / \sigma^2(\ln Y). \quad (6.c)$$

It may be noted that when the last element of  $Z$  is excluded, the remaining relative factor inequality weights

$$\sum_{j=1}^{J+1} \text{cov}[a_j Z_j, \ln Y] / \sigma^2(\ln Y)$$

sum exactly to  $R^2(\ln Y)$ .

One more bit of algebra proves useful. The ordinary correlation coefficient is related to the covariance by

$$\text{cor}[a_j Z_j, \ln Y] = \text{cov}[a_j Z_j, \ln Y] / \sigma(a_j Z_j) \sigma(\ln Y). \quad (7)$$

Combining (6.a-c) and (7), we then have:

**Result 1:** Given the income-generating function (3.a-c), let  $s_j(\ln Y)$  denote the share of the log-variance of income that is attributable to the  $j$ 'th explanatory factor and let  $R^2(\ln Y)$  be the fraction of the log-variance that is explained by all of the  $Z$ 's taken together. Then, the log-variance of income can be decomposed as

$$s_j(\ln Y) = \text{cov}[a_j Z_j, \ln Y] / \sigma^2(\ln Y) = \frac{a_j * \sigma(Z_j) * \text{cor}[Z_j, \ln Y]}{\sigma(\ln Y)} \quad (8.a)$$

where

$$\sum_{j=1}^{J+2} s_j (\ln Y) = 100\% \quad (8.b)$$

and

$$\sum_{j=1}^{J+1} s_j (\ln Y) = R^2 (\ln Y). \quad (8.c)$$

The fraction that is explained by the j'th explanatory factor,  $p_j(\ln Y)$ , is then

$$p_j (\ln Y) \equiv \frac{s_j (\ln Y)}{R^2 (\ln Y)}. \quad (8.d)$$

Equations (8.a-d) provide a full and exact decomposition of the log-variance.

However, because of problems with the log-variance (Sen, 1973; Foster and Ok, 1999), it would be nice to be able to decompose other inequality measures besides the log-variance.

This proves to be quite possible.

## **B. Extension to Other Inequality Measures**

Result 1 can be extended to other inequality measures by borrowing from a literature which at first would appear to have nothing to do with the problem at hand, namely, the literature on decomposition of inequality by additive factor components. In this literature, the i'th recipient unit's total income  $Y_i$  is expressed as the sum of its income

from each of several factor components, e.g., labor income, capital income, transfer income, etc.:

$$Y_i = \sum_k Y_{ik}. \quad (9)$$

Let  $N$  denote the total number of income recipients,  $Y \equiv (Y_1 \dots Y_N)$ , and  $Y_k \equiv (Y_{1k} \dots Y_{Nk})$ . The question asked in this literature is, what fraction of total income inequality, gauged by an inequality measure  $I(Y)$  is accounted for by labor income, by capital income, by transfer income, etc.?

Define a "relative factor inequality weight"  $s_k$  to be the percentage of income inequality that is accounted for by the  $k$ 'th factor -- for instance, how much of the inequality of total income is accounted for by the inequality of labor income. An important theorem on decomposition by additive factor components is due to Shorrocks (1982), who shows:

**Theorem (Shorrocks, 1982):** Under the six assumptions enumerated

in the appendix, the relative factor inequality weights  $s_k$  are given by

$$s_k = \text{cov}(Y_k, Y) / \sigma^2(Y) \quad (10.a)$$

such that

$$\sum_k s_k = 1 \quad (10.b)$$

for any inequality index  $I(Y)$  which is continuous and symmetric

and for which  $I(\mu, \mu, \dots, \mu) = 0$ .<sup>8</sup>

Virtually all inequality indices satisfy these conditions, including the Gini coefficient, the Atkinson index, the generalized entropy family, the coefficient of variation, and various centile measures.

Shorrocks's theorem is directly applicable to the question dealt with here, namely, using income-generating functions to account for income inequality. The standard income-generating function written in the form

$$\ln Y_{it} = a_i Z_{it} \tag{3.a}$$

has the same additive form as the equation expressing total income as the sum of the income from each component

$$Y_i = \sum_k Y_{ik}. \tag{9}$$

Note too that when the inequality of (9) is decomposed, Shorrocks obtains

$s_k = \text{cov}(Y_k, Y) / \sigma^2(Y)$  such that  $\sum_k s_k = 1$ , which has the same form as (8) with  $Y_k$

replacing  $a_j Z_j$  and  $Y$  replacing  $\ln Y$ . Now, taking advantage of this homeomorphism and applying Shorrocks's theorem, we get the following key result:

**Result 2:** Given the income-generating function (3.a-c), let an inequality index  $I(\ln Y)$  be defined on the vector of log-incomes  $\ln Y \equiv (\ln Y_1, \dots, \ln Y_N)$ .

Under the six axioms enumerated in the appendix, the decomposition of

income inequality given by

$$s_j(\ln Y) = \text{cov}[a_j Z_j, \ln Y] / \sigma^2(\ln Y) = \frac{a_j * \sigma(Z_j) * \text{cor}[Z_j, \ln Y]}{\sigma(\ln Y)} \quad (8.a)$$

where

$$\sum_{j=1}^{J+2} s_j(\ln Y) = 100\%, \quad (8.b)$$

$$\sum_{j=1}^{J+1} s_j(\ln Y) = R^2(\ln Y), \text{ and} \quad (8.c)$$

$$p_j(\ln Y) \equiv \frac{s_j(\ln Y)}{R^2(\ln Y)} \quad (8.d)$$

holds for any inequality index  $I(\ln Y_1, \dots, \ln Y_N)$  which is continuous and symmetric and for which  $I(\mu, \mu, \dots, \mu) = 0$ .

These conditions can be shown to hold for a broad class of inequality measures by the following argument.<sup>9</sup> The standard inequality measures defined on the vector of incomes  $Y = (Y_1, \dots, Y_N)$  are continuous and symmetric functions that equal zero when all income recipients receive the mean income. In such a function, substitute the identity  $e^{\ln Y_i} \equiv Y_i$  wherever  $Y_i$  occurs. The resultant inequality measure  $I(\ln Y_1, \dots, \ln Y_N)$  defined on the vector of log-incomes is also continuous and symmetric and satisfies the

property  $I(\mu, \mu, \dots, \mu) = 0$ , and therefore the factor inequality weights from Result 2 can be applied to these standard measures.

Result 2 is quite powerful. It says that as long as we agree on the log-linear model (3) and on the decomposition rules, we do not need to agree on which particular inequality measure to decompose, because we get the same percentage effect for the  $j$ 'th explanatory factor for a broad class of inequality measures applied to the logarithms of income. Included in this class are the Gini coefficient, the Atkinson index, the generalized entropy family, and various centile measures.

### **C. Different Kinds of Explanatory Variables in Practice**

We have just seen that the percentage contribution of the  $j$ 'th variable to total income inequality is given by (8). For explanatory factors that enter the income-generating function as simple variables (e.g., years of education or a dummy variable for union membership), each of the components on the right hand side of (8.a) has a straightforward interpretation. However, not all explanatory variables enter the earnings function in this way. There are three types of such variables: (1) A categorical variable entered as a string of dummy explanatory variables,<sup>10</sup> (2) An explanatory variable which has a non-linear (say, quadratic) effect,<sup>11</sup> and (3) Two or more explanatory variables which enter interactively.<sup>12</sup>

To deal with the first two of these issues in the levels decomposition, the solution is the same. Define the generic factor "industry" as the composite of the industry dummy variables  $IND_1, IND_2, \dots$  and the generic factor "experience" as the composite of EXP and

EXP<sup>2</sup> (and higher-order terms if included). Sum the s<sub>j</sub>'s for IND<sub>1</sub>, IND<sub>2</sub>, . . . to get a good measure of the overall importance of "industry," and likewise for EXP and EXP<sup>2</sup> for a measure of the importance of "experience."

Interactions pose more of a problem. Thus far, the J variables determining log-income have been assumed to enter the income-generating function additively. As long as this assumption is maintained, the model gives a factor inequality weight s<sub>j</sub> for each factor, these factor inequality weights are identical for a broad class of inequality measures, and the sum of these factor inequality weights is R<sup>2</sup>. Thus, for example, to account for inequality among a sample of working men and women using a Mincer-type human capital specification, one might run an income-generating function of the form

$$\log y = a + b_1 \text{ EDUC} + b_2 \text{ EXP} + b_3 \text{ EXPSQ} + b_4 \text{ GENDER} + e_i \quad (11)$$

and, using the results above, derive s<sub>j</sub>'s for education, experience, and gender.

One might object to the specification in (11) on the grounds that it assumes that education and experience have the same effect on income regardless of gender, whereas ample empirical research shows that this is not the case (e.g., Blau, Ferber, and Winkler, 1998). Suppose that the analyst wished to include these variables interactively, thereby allowing for the possibility that education and experience affect income differently for men and for women. One way of doing this would be to interact gender with the other variables in a single equation

$$\begin{aligned} \log y = & a + b_1 \text{ EDUC} + b_2 \text{ EXP} + b_3 \text{ EXPSQ} + b_4 \text{ GENDER} \\ & + b_5 \text{ GENDER*EDUC} + b_6 \text{ GENDER*EXP} + b_7 \text{ GENDER*EXPSQ} + e_{ii} . \end{aligned} \quad (12)$$

The problem with this is that the seven resulting  $s_j$ 's would no longer decompose neatly into education, experience, and gender components.

Another way of allowing for interactions would be to run separate income-generating functions for men and for women

$$\log y^m = a^m + b_1^m \text{ EDUC} + b_2^m \text{ EXP} + b_3^m \text{ EXPSQ} + e_i^m \quad (13.a)$$

$$\log y^f = a^f + b_1^f \text{ EDUC} + b_2^f \text{ EXP} + b_3^f \text{ EXPSQ} + e_i^f \quad (13.b)$$

and to regard inequality in the full sample as consisting of inequality among men, inequality among women, and inequality between men and women. The problems with this way of allowing for interactions are that the class of axiomatically-justified ways of cardinalizing such a decomposition remains quite broad, and further, the results are not identical for different inequality measures that might be chosen.<sup>13</sup> Thus, under this option, the agreement of results independently of the inequality measure chosen is lost.

It is up to the individual analyst to decide which is the best choice for him or her. The empirical work below adopts the first of these, i.e., equation (11).

#### **D. Comparison with Other Level Decompositions**

Decompositions in the human capital tradition have a long history dating back to the pioneering work of Mincer (1958, 1970, 1974), Becker (1964, 1967), and others. One such decomposition was suggested by Chiswick and Mincer (1972), who showed that when earnings depend on schooling (S), experience (EXP), and weeks worked (WEEKS) in the following way

$$\ln(w_i) = a + b_1 S_i + b_2 (A_i - S_i - 5) + b_3 \ln(\text{WEEKS}_i) + \varepsilon, \quad (14)$$

then income inequality as measured by the log-variance can be decomposed as

$$\begin{aligned}
 \sigma^2(\ln(w_i)) &= (b_1 - b_2) \sigma^2(S) \\
 &+ b_2^2 \sigma^2(A) \\
 &+ b_3^2 \sigma^2(\ln\text{WEEKS}) \\
 &+ [2b_2 (b_1 - b_2)] R_{a,s} \sigma(A) \sigma(S) \\
 &+ [2b_3 (b_1 - b_2)] R_{\ln\text{weeks}, s} \sigma(\ln\text{WEEKS}) \sigma(S) \\
 &+ [2b_2 b_3] R_{a,\ln\text{weeks}} \sigma(A)\sigma(\ln\text{WEEKS}) \\
 &+ \sigma^2(\varepsilon). \tag{15}
 \end{aligned}$$

The first three terms on the right hand side of (15) are the variances of schooling, age, and log-weeks weighted by the regression coefficients; the next three are interactions among the regressors; and the last is the variance of the error term. The strength of this method is that it decomposes the percentage of inequality explained by the regressors (64.8% in Chiswick and Mincer's empirical application for the United States) into components associated with schooling, experience, and weeks worked.<sup>14</sup> On the other hand, the Chiswick-Mincer method does not give "pure" effects of the regressors, it cannot handle a quadratic in experience or in other variables, and it quickly becomes unwieldy as further explanatory variables are added.

More recently, Mincer (1997) has decomposed the log-variance into four components:

- I: the variance due to schooling wage differentials
- II: the residual variance at overtaking, reflecting differentials within schooling groups
- III: the variance component due to differences in returns to post-school investments
- IV: the contribution of between-experience-group wage differentials, which reflects the steepness of the age income-generating profile.

This decomposition shares the same features as those raised at the end of the preceding paragraph.

Other variance-based decompositions have been proposed. Given two sources 1 and 2 such that  $Y = X_1 + X_2$ , we know that

$$\text{Var}(Y) = \mu_2^2 \text{var}(X_1) + \mu_1^2 \text{var}(X_2) + 2 \mu_1 \mu_2 \text{cov}(X_1, X_2).$$

$$\underbrace{\hspace{2cm}}_A \quad \underbrace{\hspace{2cm}}_B \quad \underbrace{\hspace{2cm}}_C$$

Goldberger (1970) reports Burt and Finley's (1968) suggestion to allocate  $A + B + C$  into  $A + C/2$  as the share of  $X_1$  and  $B + C/2$  as the share of  $X_2$ . Given  $Y = \sum_k Y_k$ , the variance can be decomposed so that half the value of all the interaction terms involving factor  $k$  is assigned to that factor (Shorrocks, 1982, 1999). Because the decomposition rule given in Result 2 is equivalent to the "natural" decomposition of the log-variance, that same fifty-fifty assignment of the covariance holds for the present method as well.

A number of other decompositions have appeared in the literature based on linear income-generating functions. Both the standard ANOVA model and the regression-based alternative proposed by Behrman, Knight, and Sabot (1983) give the proportion of

the log-variance of earnings explained by each independent variable. However, in neither method are the shares due to each factor derived axiomatically, as Shorrocks's  $s_j$ 's are.

Another regression-based framework is that of Morduch and Sicular (1998). Income (rather than its logarithm) is regressed on a number of explanatory variables. The main empirical conclusion from Morduch and Sicular's work is that the results vary enormously.<sup>15</sup> This is why a robust decomposition rule, derived axiomatically, may be preferable.

A different strand in the literature abandons the regression framework entirely and examines between-group and within-group inequality. For example, Cowell and Jenkins (1995) partition the population into a set of mutually exclusive and jointly exhaustive subgroups -- in their empirical application to the United States, forty-eight sex-race-age-employment status cells are used. The authors then calculate how much inequality is between sex cells, between sex-race cells, and so on and find that "not very much" of U.S. inequality is explained by population characteristics. More important for present purposes is the fact that unlike the method devised in Sections 2.A and 2.B, the relative contributions of these various characteristics (sex, race, age, and employment status) depend critically on the order in which they are introduced into the analysis. Alternatively, one might introduce the various characteristics one by one, but then the effects are gross ones not controlling for the effect of any other variable.<sup>16</sup>

Other authors have considered the role of one or a small number of explanatory variables. For example, Almeida and Barros (1991) determined how much education

contributes to the overall inequality of wages by calculating the amount by which inequality (measured by Theil's L-index) would fall if proportional transfers were made from better-educated groups to less-educated groups so that the group means were equalized. They estimate that such transfers would cut wage inequality in Brazil in half. However, because they include only a single explanatory variable, there is no way to tell whether education contributes more to inequality in this sense than do other factors. Lam and Levison (1991) adopted a similar procedure. Lam (1999) included three explanatory variables (schooling, age, and race) and found that schooling plays a very large role in explaining earnings inequality in both Korea and South Africa and in the latter case race plays a large role.

A quite different type of decomposition comes from the factor components literature. Fei, Ranis, and Kuo (1978) and Pyatt, Chen, and Fei (1980) decomposed total inequality into terms attributable to each factor component (e.g., labor income, capital income, land income). Fei, Ranis, and Kuo showed that the Gini coefficient of total income can be decomposed into a weighted sum of "pseudo-Ginis," the weights being given by the corresponding factor shares:

$$G(Y) = \sum_k \phi_k G(Y_k), \quad (16.a)$$

where

$Y$  = total income,

$Y_k$  = income from the k'th factor component,

$\phi_k \equiv \sum_i Y_{ik} / \sum_k \sum_i Y_{ik}$  is the share of income from factor k in total income,

and

$\mathcal{G}(Y_k)$  is the "pseudo-Gini coefficient" of income from factor k.<sup>17</sup>

Pyatt, Chen, and Fei then showed that the pseudo-Gini coefficient (which they call the "concentration ratio") is in turn the product of the ordinary factor Gini  $G(Y_k)$  and a "rank correlation ratio"

$$R_k = \frac{\text{cov}(Y_k, \rho)}{\text{cov}(Y_k, \rho_k)} = \frac{\text{covariance between factor income amount and total income rank}}{\text{covariance between factor income amount and factor income rank}} \quad (16.b)$$

and therefore:

$$G(Y) = \sum_k \phi_k G(Y_k) R_k . \quad (16.c)$$

Dividing (16.c) by  $G(Y)$ , one obtains

$$100\% = \sum_k \phi_k G(Y_k) R_k / G(Y) \equiv \sum_k \alpha_k, \quad (16.d)$$

the sum of the Fei-Ranis-Kuo-Pyatt-Chen relative factor inequality weights. Thus, both the Fei-Ranis-Kuo-Pyatt-Chen decomposition and the Shorrocks decomposition provide an additive decomposition of total inequality into the contribution of each income source. It should be noted that the relative factor inequality weights given by the two decompositions

( the  $\alpha_k$  in equation (16.d) and the  $s_j$  in equation (10.a) are not the same, the difference being due to the different decomposition rules used by the different authors.

### **3. Accounting for Differences in Income Inequality**

Section 2 established a methodology to account for the level of income inequality in a particular country at a particular time. In this section, a method is proposed to account for differences in income inequality between one country, group, or time and another.<sup>18</sup> Specifically, we ask: How much of the difference in inequality between one country/group/time and another is attributable to each income determinant? Which is relatively more important in accounting for these differences: differences in education, in experience, in gender, etc?

#### **A. The Relative Importance of Different Income Determinants in Explaining Inequality Differences**

Result 2 established that the  $j$ 'th factor's percentage contribution to the level of inequality is the same for a broad class of inequality measures. This leads one to ask, are the percentage contributions to the changes in inequality similarly independent of how inequality is measured? The answer is readily seen to be "no": the amount by which inequality rose or fell -- and perhaps even whether inequality rose or fell -- depends on how inequality is measured. Clearly, the answer to the changes question must be index-specific.

For any given inequality measure  $I(\cdot)$ , we may write the change in inequality in terms of each period's factor inequality weight and each period's inequality level as

$$I(\cdot)_2 - I(\cdot)_1 = \sum_j [s_{j,2} * I(\cdot)_2 - s_{j,1} * I(\cdot)_1]. \quad (17.a)$$

Define the contribution of factor  $j$  to the change in inequality for an arbitrary inequality measure  $I(\cdot)$  as

$$\Pi_j(I(\cdot)) \equiv [s_{j,2} * I(\cdot)_2 - s_{j,1} * I(\cdot)_1] / [I(\cdot)_2 - I(\cdot)_1]. \quad (17.b)$$

From this, we may derive

$$100\% = \frac{\sum_j [s_{j,2} * I(\cdot)_2 - s_{j,1} * I(\cdot)_1]}{I(\cdot)_2 - I(\cdot)_1} = \sum_j \Pi_j(I(\cdot)), \quad (17.c)$$

the  $\Pi_j(I(\cdot))$ 's denoting the contribution of the  $j$ 'th explanatory factor to the change in inequality measured by inequality index  $I(\cdot)$ . Thus:

**Result 3**: The contribution of the  $j$ 'th factor to the change in a particular inequality measure between country/group/time 1 and country/group/time 2 is given by

$$\Pi_j(I(\cdot)) = [s_{j,2} * I(\cdot)_2 - s_{j,1} * I(\cdot)_1] / [I(\cdot)_2 - I(\cdot)_1]. \quad (17.b)$$

Writing  $\Pi_j$  as a function of  $I(\cdot)$  makes explicit that the explanatory contribution of the  $j$ 'th factor depends on the inequality measure used. It is an empirical question whether the

choice of inequality measure makes a large difference or a small one in any particular context.

## **B. Decomposing Differences in the $s_j$ 's**

Next, let us consider how to account for the sources of changing contributions of the various factors explaining income inequality. If the same income-generating functions have been run for two samples at different dates and the  $s_j$ 's given by (8.a) are found to differ, one may ask, "why"? To what extent is the change in any given  $s_j$  due to differences between the regression coefficients in the two years? To differences in the inequality of the explanatory variable? To differences in the covariance or the correlation between the explanatory variable and income?

For infinitesimal changes, an exact decomposition of the difference in any given  $s_j$  can be obtained by logarithmically differentiating (8.a). This produces:

$$s_j(\hat{\ln Y}) = \hat{a}_j + \sigma(\hat{Z}_j) + cor[\hat{Z}_j, \ln Y] - \sigma(\hat{\ln Y}), \quad (18)$$

the  $\hat{\phantom{x}}$  over the variable indicating a percentage rate of growth. This equation proves to be directly useful in this form. In real-world applications, the changes in each component are non-infinitesimal. Dividing through by  $pctchn(s_j(\ln Y))$ , the change in  $s_j$  may then be approximated by

$$1 \approx pctchn(a_j) / pctchn(s_j(\ln Y)) + pctchn[\sigma(Z_j)] / pctchn(s_j(\ln Y))$$

$$+ \text{pctchnng}[\text{cor}[Z_j, \ln Y]] / \text{pctchnng}(s_j(\ln Y)) - \text{pctchnng}[\sigma(\ln Y)] / \text{pctchnng}(s_j(\ln Y)).^{19}$$

(19)

An objection can be raised to the right-most decomposition of levels in (8.a) and the consequent decomposition of changes in (19), which is that  $a_j$  and  $\text{cor}[Z_j, \ln Y]$  are both functions of  $\text{cov}[Z_j, \ln Y]$ , so that one cannot be varied without the other.<sup>20</sup> This objection can be overcome by making a further approximation. If the  $j$ 'th income-determining factor were orthogonal to the other income-determining factors, that determinant's factor inequality weight would equal

$$s_j(\ln Y) = \frac{a_j^2 * \sigma^2(Z_j)}{\sigma^2(\ln Y)}.$$

(20)

The changes over time would then decompose approximately as

$$1 \approx 2 * \text{pctchnng}(a_j) / \text{pctchnng}(s_j(\ln Y)) + 2 * \text{pctchnng}[\sigma(Z_j)] / \text{pctchnng}(s_j(\ln Y)) - 2 * \text{pctchnng}[\sigma(\ln Y)] / \text{pctchnng}(s_j(\ln Y)).$$

(21)

("Approximately" for two reasons: (i) real-world changes are not infinitesimal, and (ii) the  $j$ 'th regressor is typically not orthogonal to the other regressors.) On the other hand, the advantage of the decomposition in (21) over that in (19) is that it says that the  $j$ 'th regressor in the income-generating function contributes more to explaining an observed increase in inequality (a) the larger is the increase in the regression coefficient of that variable, and (b) the larger is the increase in the inequality of that variable as measured by the standard deviation -- both intuitively appealing results. In the case of falling inequality, (21) says that the  $j$ 'th regressor contributes more to the decrease in inequality

(a) the larger is the decrease in the regression coefficient on that factor and (b) the larger is the decrease in the standard deviation of that factor.

Using these alternative decompositions, we then have:

**Result 4:** The change in the j'th explanatory factor's relative factor inequality weight can be expressed as

$$1 \approx \text{pctchn}(a_j) / \text{pctchn}(s_j(\ln Y)) + \text{pctchn}[\sigma(Z_j)] / \text{pctchn}(s_j(\ln Y)) \quad (19)$$

$$+ \text{pctchn}[\text{cor}[Z_j, \ln Y]] / \text{pctchn}(s_j(\ln Y)) - \text{pctchn}[\sigma(\ln Y)] / \text{pctchn}(s_j(\ln Y))$$

or as

$$1 \approx 2 * \text{pctchn}(a_j) / \text{pctchn}(s_j(\ln Y)) + 2 * \text{pctchn}[\sigma(Z_j)] / \text{pctchn}(s_j(\ln Y)) \quad (21)$$

$$- 2 * \text{pctchn}[\sigma(\ln Y)] / \text{pctchn}(s_j(\ln Y)) .$$

### **C. Decomposing Total Inequality Into Price Effects and Quantity Effects**

Another decomposition is possible if one is willing to choose the log-variance (i.e., the variance of the logarithms of income) as the measure of income inequality; this decomposition is due to Yun (2002). In the work of Juhn, Murphy, and Pierce (described below), the price effect of a variable in accounting for the change in inequality between a base income distribution "1" and a comparison income distribution "2" is defined as the difference between the inequality of distribution 2 and the inequality of an auxiliary distribution which uses the prices of distribution 1 and the quantities and residuals of distribution 2. For the i'th individual, income in the auxiliary distribution is given by

$$\ln Y_{i,aux} = \sum_j a_{j1} Z_{ij2} \quad (22)$$

From (8), the variance of  $\ln Y_{aux}$  can be composed as

$$\sigma^2(\ln Y_{aux}) = \sum_j a_{j1} \sigma(Z_{j2}) \text{cor}(Z_{j2}, \ln Y_{aux}) \sigma(\ln Y_{aux}). \quad (23)$$

Using the auxiliary distribution, we have that the difference in inequality between distributions 1 and 2 can be expressed as

$$I_2 - I_1 = (I_2 - I_{aux}) + (I_{aux} - I_1),$$

which, for the log-variance, decomposes as

$$\begin{aligned} & \sigma^2(\ln Y_2) - \sigma^2(\ln Y_1) \\ &= \sum_j a_{j2} \sigma(Z_{j2}) \text{cor}(Z_{j2}, \ln Y_2) \sigma(\ln Y_2) \\ & \quad - \sum_j a_{j1} \sigma(Z_{j2}) \text{cor}(Z_{j2}, \ln Y_{aux}) \sigma(\ln Y_{aux}) \\ & \quad + \sum_j a_{j1} \sigma(Z_{j2}) \text{cor}(Z_{j2}, \ln Y_{aux}) \sigma(\ln Y_{aux}) \\ & \quad - \sum_j a_{j1} \sigma(Z_{j1}) \text{cor}(Z_{j1}, \ln Y_1) \sigma(\ln Y_1). \end{aligned} \quad (24)$$

Upon regrouping, we obtain

**Result 5.** The change in inequality between two distributions 1 and 2 can be decomposed as

$$\begin{aligned} & \sigma^2(\ln Y_2) - \sigma^2(\ln Y_1) \\ &= \sum_j [a_{j2} \sigma(Z_{j2}) \text{cor}(Z_{j2}, \ln Y_2) \sigma(\ln Y_2) - a_{j1} \sigma(Z_{j2}) \text{cor}(Z_{j2}, \ln Y_{aux}) \sigma(\ln Y_{aux})] \\ & \quad + \sum_j [a_{j1} \sigma(Z_{j2}) \text{cor}(Z_{j2}, \ln Y_{aux}) \sigma(\ln Y_{aux}) - a_{j1} \sigma(Z_{j1}) \text{cor}(Z_{j1}, \ln Y_1) \sigma(\ln Y_1)]. \end{aligned}$$

(25)

The variables in (25) have a clear interpretation: in Juhn, Murphy, and Pierce's terminology, each term in the first summation is the price effect of the j'th variable, while each term in the second summation is the quantity effect of the j'th variable.

Finally, in parallel with the decompositions in (19) and (21), we may want to know what fraction of the j'th variable's factor inequality weight is attributable to the price effect of the j'th variable and what fraction to its quantity effect. Taking the terms for the j'th variable and dividing by the change in that variable's factor inequality weight, we obtain:

**Result 6:** The change in the j'th explanatory factor's relative factor inequality weight can be expressed as

$$1 = \frac{[a_{j2}\sigma(Z_{j2})\text{cor}(Z_{j2}, \ln Y_2)\sigma(\ln Y_2) - a_{j1}\sigma(Z_{j2})\text{cor}(Z_{j2}, \ln Y_{aux})\sigma(\ln Y_{aux})]}{s_j(\ln Y_2) - s_j(\ln Y_1)} + \frac{[a_{j1}\sigma(Z_{j2})\text{cor}(Z_{j2}, \ln Y_{aux})\sigma(\ln Y_{aux}) - a_{j1}\sigma(Z_{j1})\text{cor}(Z_{j1}, \ln Y_1)\sigma(\ln Y_1)]}{s_j(\ln Y_2) - s_j(\ln Y_1)}, \quad (26)$$

where the first term is the percentage contribution of the price effect and the second is the percentage contribution of the quantity effect.

This completes the presentation of the decompositions for analyzing differences in the income distributions between one year/group/place and another. Before turning to empirical applications of these three methods, let us now compare the decompositions

given by (19), (21), and (26) with other decompositions that have been carried out by others.

#### **D. Comparison with Other Difference Decompositions**

Decomposing differences in income inequality in the way described in Results 3, 4, and 6 offers several advantages compared with other decompositions of inequality differences that have been suggested in the literature.

Using an income-generating function framework, Juhn, Murphy, and Pierce (1991, 1993) and followers (e.g., Blau and Kahn, 1997; Robbins and Gindling, 1999) have decomposed the change over time in quantile differentials (90-50, 90-10, and 50-10) into components due to changes in observed quantities, components due to changes in observed prices, and a residual (termed "unobserved prices and quantities"). One advantage of the decomposition proposed here is that it uses a more comprehensive measure of inequality than the 90-50, 90-10, or 50-10 differentials, and in fact much of the decomposition analysis can be done entirely non-parametrically. Another advantage is that the factor inequality weights derived here (the  $s_j$ 's) measure the relative importance of each particular explanatory factor rather than the changes in prices or the changes in quantities taken as a group.

Another literature has decomposed the difference in wage inequality between unionized and non-unionized workers. Freeman (1980) showed that given two income-generating functions

$$\ln(Y_{i1}) = a_1 + \sum_{j=1}^J b_{j1} x_{ij1} + \varepsilon_{i1} \quad (2.a)$$

and

$$\ln(Y_{i2}) = a_2 + \sum_{j=1}^J b_{j2} x_{ij2} + \varepsilon_{i2}, \quad (2.b)$$

the extent to which  $\text{var}(\ln(Y_{i1}))$  differs from  $\text{var}(\ln(Y_{i2}))$  as a result of differences in the characteristics in the samples can be gauged by

$$\sum_{j=1}^J (b_j)^2 [\sigma^2(x_{j1}) - \sigma^2(x_{j2})] + \sum_{j=1}^J \sum_{j'=1}^J b_j b_{j'} [\sigma(x_{j1} x_{j'1}) - \sigma(x_{j2} x_{j'2})] \quad (27)$$

where  $\sigma^2(x_{j1})$  is the variance in characteristic  $j$  in group 1,  $\sigma(x_{j1} x_{j'1})$  is the covariance in characteristics  $j$  and  $j'$  among members of group 1, and  $\sigma^2(x_{j2})$  and  $\sigma(x_{j2} x_{j'2})$  are the corresponding terms in group 2. But as Freeman notes, (26) is only an "approximate standardization for differences in characteristics," because all second- and higher-order covariance terms are omitted. The decomposition using (8) gives an exact standardization.

A more recent follow-up literature has estimated the effect of declining unionization rates in the United States on the log-variance of wages (Freeman, 1993; Card, 1996). Both these studies estimate the effect of the change in unionization rates on the log-variance of wages assuming that the union-nonunion wage gap and the within-sector effect of unions on the log-variance of wages are unchanged. If these factors do change, then the Freeman and Card methodologies do not give an answer. By contrast, the method presented above is multivariate and applies in such situations.

Another regression-based approach is to be found in two papers by Bourguignon and co-authors (Bourguignon and Martinez, 1997; Bourguignon, Fournier, and Gurgand, 1998). The essence of their procedure is to run two regressions for a base year 1 and a final year 2 and then to decompose the changes into price, quantity, and residual effects. Given the basic wage equations

$$\ln w_{1i} = x_{1i} b_1 + u_{1i} \quad (28)$$

and  $\ln w_{2j} = x_{2j} b_2 + u_{2j}$

for the two years, the decomposition equations are

$$W_2 - W_1 = \Delta_b + \Delta X + \Delta_\sigma, \quad (29)$$

where

$$\Delta_b = X_1 (b_2 - b_1); \Delta X = b_1 (X_2 - X_1); \Delta_\sigma = (U_2 - U_1). \quad (30)$$

(The capital letters in (28) and (30) signify vectors.) With adjustment for participation or not and employed or not, the model becomes

$$\ln w_{ki} = P_{ki}(x_{ki}, y_{ki}) * E_{ki}(x_{ki}, y_{ki}) * (x_{ki} b_k + u_{ki}) \quad (31)$$

with  $P_{ki}(x_{ki}, y_{ki}) = 1/0$  as  $x_{ki} a_k + y_{ki} c_k + v_{ki} \geq < 0$

and  $E_{ki}(x_{ki}, y_{ki}) = 1/0$  as  $x_{ki} d_k + y_{ki} f_k + t_{ki} \geq < 0$

and an analogous decomposition is performed.

The Bourguignon method gives a thorough accounting of the routes by which a change in an explanatory factor affects income inequality. However, implementing that method requires heavy econometrics. By contrast, the method derived above is easier to apply, but it produces only an incomplete decomposition. What is interesting empirically is that both methods have been used to understand changing income inequality in

Taiwan, and they both produced the conclusion that the major factors affecting income inequality there were an increased coefficient on education, which raised income inequality, and a reduction in the inequality of years of education, which lowered income inequality (Bourguignon, Fournier, and Gurgand, 1998; Fields and Mitchell, 1999).

Then, there is the recent and comprehensive method of DiNardo, Fortin, and Lemieux (1996). These authors gauge the effect of various labor market changes on the density function of wages in the United States between 1979 and 1988. Once they estimate the effect of a given change on the entire density function, they then calculate the implied effect on various inequality measures including the Gini coefficient, the Theil index, and various percentile differentials (e.g., 90-10).

Their method is most easily understood if we adopt the following notation: <sup>21</sup>

A = actual wage distribution in 1988

B = 1988 wage distribution adjusted for 1979 minimum wage

C = 1988 wage distribution adjusted for 1979 minimum wage and 1979 unionization

D = 1988 wage distribution adjusted for 1979 minimum wage, 1979 unionization, and

1979 other attributes ( which include experience, schooling, race, full-time/part-time, and dummy variables for SMSA, occupation, and industry)

E = 1988 wage distribution adjusted for 1979 minimum wage, 1979 unionization, 1979 other attributes, and 1979 supply and demand

F = 1988 wage distribution adjusted for everything including a residual, which is then the actual wage distribution in 1979.

Their key equation takes the form

$$A - F = [A - B] + [B - C] + [C - D] + [D - E] + [E - F], \quad (32)$$

where

$A - B =$  1988 wage distribution adjusted for 1979 minimum wage = “effect” of minimum wage<sup>22</sup>

$B - C =$  B adjusted for change in unionization rate = “effect” of unionization

$C - D =$  C adjusted for changes in other attributes = “effect” of other attributes

$D - E =$  D adjusted for supply and demand = “effect” of supply and demand

$E - F =$  E adjusted for residual = “effect” of residual.

The principal advantages of the DiNardo-Fortin-Lemieux method compared with the decomposition procedure described in Results 3 and 4 are that it estimates the effect of a given income determinant on the entire wage distribution and it does not rely on a particular functional form. However, the main disadvantage of their method is that each "effect" depends on the order in which the adjustment is done.

On the other hand, the method derived here has advantages of its own. Because it relies on a regression framework, it expresses inequality levels and inequality changes as functions of the very income determinants that economists are accustomed to using. Also, in explaining income levels, it assigns the same weights to each income determinant regardless of the inequality measure used. And as a practical matter, the required calculations in the present method are easier to make. It is left to the reader to weigh the two methods' respective pros and cons.

Finally, mention should be made of two other decompositions with quite different purposes. Moffitt and Gottschalk (1995) decompose changes in inequality over time into

two components, one representing the variance in the permanent component of income and the other the variance in the transitory component. Layard and Zabalza (1979) decompose the inequality of family income into the variances and covariances of incomes of family members. Neither of these is directly relevant to the problem considered here.

#### **4. Empirical Application: Analyzing the Sources of Rising Earnings Inequality in the United States**

Official publications and academic studies have shown a substantial increase in income inequality in the United States over the last twenty years (Katz and Autor, 1999; Economic Report of the President, 1999; Levy, 1999; Forster and Pellizzari, 2000). In this section, the methods developed in Sections 2 and 3 are used to quantify the contributions of various factors in accounting for the amount of labor earnings inequality at a point in time (the "levels" question) and also the increase in inequality of labor earnings (the "differences" question).

Data for this analysis come from the Annual Demographic Surveys (March supplements) to the 1980 and 2000 U.S. Current Population Surveys. The March supplements contain respondents' reports on labor earnings in the preceding year. The 1980 CPS was chosen as the starting point, because it has been the base year for a number of important empirical studies (Katz and Murphy, 1992; Bound and Johnson, 1992; DiNardo, Fortin, and Lemieux, 1996) and because earnings inequality had not yet started

rising in the United States at that time. The 2000 CPS was chosen as comparison year so as to be able to speak about changes in inequality over a twenty year period.

Following Katz and Murphy (1992), Juhn, Murphy, and Pierce (1993), Blau (1998), and Katz and Autor (1999), the sample and variables were defined as follows. The sample consists of men and women who were full-year, full-time wage and salaried employees,

18-64 years of age. The dependent variable used in the analysis is weekly earnings, measured in logs. The explanatory variables are gender (2 categories), race (2 categories), education (4 categories, also entered continuously – see below), potential experience and its square, occupation (3 categories), industry (3 categories), and geographic region (4 categories). Descriptive statistics on these variables are presented in Table 1.

--- Insert Table 1 around here ---

In this data set, the distribution of labor earnings became unambiguously more unequal. Figure 1 displays the Lorenz curves, showing a clear Lorenz-worsening and therefore an increase in inequality for a broad range of inequality measures. The most commonly-used summary statistic of inequality, the Gini coefficient, rose from 0.274 in 1979 to 0.338 in 1999 among this sample of workers. Another commonly-used inequality measure, the variance of the logarithms of income, increased too, from 0.262 to 0.380. Increases in inequality of this magnitude are large, both by the standards of changes that

typically take place within countries and by the standards of international differences in inequality at a point in time (Atkinson, 1997; Forster and Pellizzari, 2000).

--- Insert Figure 1 around here ---

The first step in the decomposition analysis is to run the earnings functions (1) and (2). Log-earnings is a linear function of gender, race, potential experience and its square, four schooling categories, three occupational categories, three industry categories, and four region categories. The empirical results are given in Table 2. We see that in both years, all variables included in the regression are statistically significant at conventional levels, and together they explain 41.5% of the variance of log-earnings in 1979 and 38.3% of the variance in 1999.

--- Insert Table 2 around here ---

The levels question is: Of these statistically significant variables, which are how important in accounting for the levels of inequality in 1979 and 1999? The answer is given in columns (1) and (2) of Table 3. In 1999, after the residual, schooling was the most important variable, with a factor inequality weight of 16.1%. Other variables with sizeable shares were occupation (9.1%), experience (6.6%), and gender (5.7%). Three other variables had shares that were effectively zero – region (0.5%), race (0.4%), and industry (0.0%). What we see, then, is that although all of these explanatory factors were

statistically significant determinants of earnings levels, their importance differs enormously: schooling was about twice as important as each of the next three closest variables and orders of magnitude higher than the three least important variables. These differences in relative importance could not have been seen from standard regression output alone.<sup>23</sup>

--- Insert Table 3 around here ---

Turning our attention now to the question of how much of the increase in earnings inequality was due to each of these factors ("the differences question"), columns (3) and (4) of Table 3 give the answer using equation (17.b) in Result 3. The decompositions of both inequality measures show that the largest share of the increase in earnings inequality was accounted for by an increase in residual inequality. Previously, it was known that earnings inequality had increased within education/experience/ . . . cells (Katz and Murphy, 1992; Murphy and Welch, 1992; Juhn, Murphy, and Pierce, 1993; Welch, 1999) but the relative weight of this factor vis-à-vis other factors was not known. Looking at "real" variables, we find here too that the differences in their explanatory contributions are enormous. Schooling is the largest such variable, accounting for 56% of the increase in the Gini coefficient and 34% of the increase in the log-variance. Occupation was half as important as schooling, accounting for 28% of the increase in the Gini coefficient and 18% of the increase in the log-variance. Four other variables – race, experience, industry, and region -- contributed essentially nothing. Finally, gender had a sizeable effect but it does

not contribute to the explanation; because gender changes were in the equalizing direction but inequality increased, gender's weight is strongly negative. In sum, some of the variables that were found in columns (1) and (2) to be statistically significant determinants of earnings levels are economically insignificant in accounting for changes in earnings inequality.

Looking more deeply into the role of the single most important variable, schooling, we may ask to what extent schooling's contribution to rising earnings inequality was due to increased dispersion of earnings between workers with different educational attainments, to what extent to increased inequality of years of schooling, and to what extent to other factors. So that the inequality of years of schooling could be calculated in answering this question, schooling in categories was replaced by a continuous schooling variable, entered linearly. The regression results appear in Appendix Table 1 and the first decomposition results in Appendix Table 2. These results show a rise in the coefficient on schooling (sometimes called a "rate of return"), which has been found in many, many prior studies. They also show reduced inequality in years of schooling, which as well has been found in past work. Based on these findings, we can further decompose the effect of schooling to understand why it contributed what it did, using equations (19), (21), and (26) in Results 4 and 6. The results, presented in Table 4, show that using all three alternative methods, schooling contributed to rising inequality entirely because of a rising coefficients effect and not at all because of an increase in inequality of years of schooling.

--- Insert Table 4 around here ---

The final step in the U.S. analysis is to disaggregate by gender. This is done because the gender variable makes a negative contribution to explaining rising inequality, and so we want to know whether the same factors that are important or unimportant for the two genders taken together are similarly important or unimportant when the two are considered separately. The results are shown in Table 5.

--- Insert Table 5 around here ---

The top part of the table shows that earnings inequality increased for both women and men. To understand why, earnings equations were run within gender (using schooling in years) and factor inequality weights and the contributions of each factor to changes in inequality were calculated.

For women, after the residual, the big variables accounting for earnings inequality in each year are schooling and occupation. These are also the biggest variables explaining the increase in earnings inequality, with schooling exhibiting about twice as large an effect as occupation. Lastly, for women, schooling contributed to rising inequality entirely because of increased differences in earnings across schooling levels and not at all because of inequality of years of schooling; in fact, years of schooling became distributed slightly more equally for women during that twenty year period.

For men, schooling was also a leading factor accounting for rising earnings inequality in both years. Unlike women, for men, experience was as important a factor in

1979 as schooling was, but by 1999, its relative contribution had fallen in half.

Occupation's role increased from 1979 to 1999 and came to equal the role of experience in the latter year. As with women, in explaining the increase in earnings inequality for men, schooling had the largest effect, followed by occupation; experience and other variables explained virtually nothing. Finally, for men as for women, schooling's contribution to rising inequality is seen to be a coefficients effect and not an inequality-of-schooling effect.

In sum, for the empirical study of the United States, we have learned that earnings inequality increased overall and for women and men separately; that seven variables were statistically significant determinants of earnings (gender, race, potential experience, schooling, occupation, industry, and region); that despite all being statistically significant, their contribution to rising inequality differed enormously, with schooling being far and away the most important variable and many other variables (race, experience, gender, industry, and region) having no role to play at all; and that schooling's effect was entirely a coefficients effect and not at all due to rising inequality of schooling.

## **5. Conclusion**

This paper has presented a methodology to account for a) income inequality levels in a given country, group, or time period and b) differences in income inequality between one country and other, between one group and another, or between one time period and another. To sum up what should be done:

- For a log-income-based levels calculation, run a standard semi-logarithmic income-generating function (3) for a particular country, group, or date. Using equations (8.a-d) in Result 2, calculate the relative factor inequality weights  $s_j(\ln Y)$  and the corresponding percentage contributions  $p_j(\ln Y)$  for each explanatory factor. If you accept the decomposition rules in the appendix, you get the same  $s_j(\ln Y)$ 's and  $p_j(\ln Y)$ 's for virtually any inequality measure calculated on the vector of log-incomes.
- To account for the role of a given income determinant in explaining the change in inequality based on a particular inequality measure  $I(\cdot)$ , you can use equation (17.b) in Result 3 to gauge the proportion of the rise or fall in inequality according to that measure that is accounted for by each explanatory factor.
- Finally, to explain why each explanatory factor contributed to an increase or decrease in income inequality between one country/group/date and another, you can use equations (19), (21), and (26) in Results 4 and 6 to get a decomposition into a coefficients effect, a standard deviation effect, and a correlation effect.

As an application of this approach, these methods were then used to analyze labor earnings inequality in the United States in 1979 and 1999 and the increase in labor earnings inequality between those two years. Explanatory variables included gender, race, potential experience, schooling, occupation, industry, and region. Although all variables were found to be statistically significant determinants of earnings in both years, the decomposition analysis revealed enormous differences in their explanatory power. In

explaining the levels of inequality, schooling exhibited the largest explanatory power, followed by occupation, experience, and gender; the three remaining variables – region, race, and industry – had no appreciable effect at all. Then, in explaining the increase in inequality, schooling was again the single most important variable, but only one other variable (occupation) has any appreciable role to play; gender worked in the equalizing direction and all other variables contributed essentially zero explanatory power. All of schooling's effect was a coefficients effect and none an inequality-of-schooling effect. All of these results hold when women and men are analyzed separately.

\* \* \*

Before ending, one final point bears repetition. Although this entire paper has been cast in terms of income inequality, this methodology can be used to apply regression analysis to the decomposition of anything. The usefulness of the method is limited only by the meaningfulness of the regression to which it is applied.

--- Insert Appendix Tables 1 and 2 around here following the text ---

## APPENDIX

### Conditions on the Decomposition.

In the text, Shorrocks's theorem makes reference to six conditions on the decomposition itself. Let  $Y^{ik}$  denote the income of the  $i$ 'th income recipient from factor  $k$ ,  $Y^k = (Y^{i1} \dots Y^{iK})$  be the vector of incomes from the  $k$ 'th factor,  $Y^i = \sum_k Y^{ik}$  be the  $i$ 'th recipient's total income,  $N$  be the total number of income recipients, and  $K$  be the total number of factor income components. Let  $I(Y)$  be an inequality measure defined on the space of total incomes  $Y = (Y^1 Y^2 \dots Y^N)$  and let  $S_k = S_k(Y^1, \dots, Y^K; K)$  be the amount of inequality accounted for by each of the  $K$  components. Using this notation, Shorrocks's six conditions may be expressed thus:

Condition 1: (Number of Components) The inequality measure  $I(Y)$  is to be divided into  $K$  components, one for each income factor, denoted  $S_k(Y^1, \dots, Y^K; K)$ .

Condition 2: (a) (Continuity) Each  $S_k$  is continuous in  $Y^k$ . (b) (Symmetric Treatment of Factors) If  $\pi_1, \dots, \pi_k$  is any permutation of  $1, \dots, K$ ,  $S_k(Y^1, \dots, Y^K; K) = S_{\pi k}(Y^{\pi_1}, \dots, Y^{\pi_k}; K)$ .

Condition 3: (Independence of the Level of Disaggregation) The amount of inequality accounted for by any one factor  $S_k$  does not depend on how the other factors are grouped.

Condition 4: (Consistent Decomposition) The contributions  $S_k$  sum to the overall amount of inequality, viz.,

$$\sum_k S_k(Y^1, \dots, Y^K; K) = I(Y).$$

Condition 5: (a) (Population Symmetry) If  $P$  is any  $n \times n$  permutation matrix,  $S(Y^k P, Y P) = S(Y^k, Y)$ ; (b) (Normalization for Equal Factor Distribution) If all income recipients have the same value for the  $k$ 'th factor, then the share of inequality accounted for by that factor  $S(\mu_k e, Y) = 0$  for all  $\mu_k$ .

Condition 6: (Two Factor Symmetry) Suppose the distribution of factor 2 incomes  $Y^2$  is simply a permutation of that for factor 1,  $Y^1$ . Then if those were the only two sources of income,  $Y^1$  and  $Y^2$  should receive the same value in the decomposition. Thus, for all permutation matrices  $P$ ,  $S(Y^1, Y^1 + Y^1 P) = S(Y^1 P, Y^1 + Y^1 P)$ .

These six conditions generate the factor inequality weights  $s_k$  given in the text by

$$s_k = \text{cov}(Y_k, Y) / \sigma^2(Y) \quad (10.a)$$

such that

$$\sum_k s_k = 1. \quad (10.b)$$

## **ACKNOWLEDGMENTS**

This work was financed in part by a grant from the Organisation for Economic Co-Operation and Development. Thanks are gratefully extended to Jesse Leary and Paola Valenti for invaluable programming assistance, to Bob Hutchens, George Jakubson, and Larry Kahn for many helpful discussions during the preparation of this paper, and to François Bourguignon, Leonard Cheng, Arthur Goldberger, Larry Katz, Jacob Mincer, Jonathan Morduch, Efe Ok, Solomon Polachek, and Grace Tsiang for useful comments and suggestions. Earlier versions of this paper were presented at Cornell, Yale, Harvard, Columbia, the City University of New York, Les Facultés Universitaires Notre-Dame de la Paix, and the American Economic Association annual meetings.

## <sup>1</sup>NOTES

<sup>1</sup> The literature uses several different income concepts: "income" denotes the recipient unit's income from all sources, "earnings" denotes income from employment or self-employment, and "wages" denotes earnings per hour. It also uses several different recipient units including families, households, and individuals. To avoid having to refer repeatedly to income/earnings/wages among households/families/ individuals/workers, the following discussion is cast in terms of "incomes" among "individuals" except for those empirical studies that specifically used something else.

<sup>2</sup> The terminology "income-generating function" is used in place of "earnings function" or "wage equation," because the method is general enough to allow for non-labor income to be included along with labor income in the regression if the analyst so chooses.

<sup>3</sup> What follows is a "decomposition" in the sense that the overall inequality in a population is broken down into a number of components such that the whole is equal to the sum of its parts. The term "decomposition" has been used in this sense in many types of income distribution studies including the literature on inequality decomposition by factor components (e.g., Fei, Ranis, and Kuo, 1978; Pyatt, Chen, and Fei, 1980; and Shorrocks, 1982) and the literature decomposing differences in mean incomes between

groups (Oaxaca, 1973; Blinder, 1974; Oaxaca and Ransom, 1994). However, the term "decomposition" has also been used in a more restrictive sense by Bourguignon (1979), who defines an income inequality measure to be decomposable when the total inequality of a population can be broken down into a weighted average of (i) the inequality existing within subgroups of the population using that same inequality measure and (ii) the inequality existing between the subgroups. In what follows, the term "decomposition" is used in the less restrictive sense, whereby the total inequality in a population is expressed as the sum of a number of terms, each corresponding to an explanatory variable in the income-generating equation.

<sup>4</sup> Since this paper has been circulated in working paper form, the decompositions derived here have been used in a number of studies including works by Arcos (1996), Fields et al. (1998), Sánchez and Núñez (1998), Fields and Mitchell (1999), Ravallion and Chen (1999), Fields and Yoo (2000), Contreras (2000), Andersen (2000), Redmond and Kattuman (2001), Gindling and Trejos (2001), Heltberg (2001), and Yun (2002).

<sup>5</sup> Note that the latter question is why one income distribution is more equal than another, not why one has a higher mean than another. The latter is the question addressed in the Blinder-Oaxaca types of decompositions.

<sup>6</sup> Throughout this paper, a single regressor in the income-generating equation is called a "variable." Sometimes, there are natural groupings of "variables" into "factors." So for example, EXP is a variable, EXPSQ is another variable, and the two together constitute the factor "experience."

<sup>7</sup> After this paper had been circulating in working paper form for some time, Arthur Goldberger brought to my attention a passage in his 1964 book (pp. 197-200) in which he stated without proof or axiomatic justification that in the standard linear model

$Y = X\beta + \varepsilon$ , the total residual sum of squares could be decomposed as

$T \text{ SSR} = b_1 m_{1y} + \dots + b_K m_{Ky}$ , where  $m_{jy} = T \sum (x_j - \bar{x}_j)(y - \bar{y})$ . This produces the so-

called "separate determination coefficient"  $d_{yj}^2 = b_j \frac{m_{jy}}{m_{yy}}$  which, after appropriate

substitutions, can be shown to equal  $b_j \frac{\sum (x - \bar{x})(y - \bar{y})}{\sum (y - \bar{y})^2} = b_j \frac{\text{cov}(x, y)}{\text{var}(y)}$ . This is precisely (6.c)

with income as the dependent variable in place of log-income. Goldberger notes that  $d_{yj}^2$  is "not widely used." I myself have never seen it used in the income inequality literature.

<sup>8</sup>  $\mu$  is the mean of total income.

<sup>9</sup> I thank an anonymous referee for this insight.

<sup>10</sup> This may arise for a variable which is inherently categorical (e.g., industry, occupation) or for a continuous variable which is censored (e.g., years of education, where all that is known is the highest level attended or completed but not the number of years).

<sup>11</sup> Experience and experience squared are commonly included in earnings functions.

<sup>12</sup> Earnings functions often contain a unionization variable interacted with years of education or experience.

<sup>13</sup> See Cowell and Jenkins (1995) for a comprehensive discussion of the issues involved.

<sup>14</sup> An even higher R-squared was found by Plotnick (1982) when log-earnings was used as the dependent variable in place of log-income and when the percentage of male workers in unions was added to the equation.

<sup>15</sup> They find that in Zouping Country (China) in 1993, the decompositions of different inequality measures produce an education effect ranging from plus 174% to minus 30% and village effects ranging from plus 125% to minus 467%.

<sup>16</sup> See also Ahuja et al.'s (1997) applications of the Cowell-Jenkins method to China and Thailand.

<sup>17</sup> The pseudo-Gini coefficient of a factor component is the Gini coefficient that is obtained if income recipients are arrayed in increasing order of total income rather than in increasing order of income from that factor.

<sup>18</sup> Though the presentation in the text is in terms of changes in income inequality over time, the same methodology can be used to account for inequality differences between one country and another or between one group and another within a country. The reader is reminded that the question here is what accounts for differences in inequality between one time/country/group and another, not what accounts for differences in means.

<sup>19</sup> This expression has the following feature. Suppose that, holding the distribution of all the  $Z_j$ s constant, all incomes were to change by the same non-zero scalar multiple. Then all terms on the right hand side of (19) would be zero, thereby satisfying the intuitively appealing adding-up constraint that the share of inequality accounted for by the  $j$ 'th explanatory factor is unchanged in such a case.

<sup>20</sup> In the case of a single regressor,  $a_j = \text{cov}(X, \ln Y) / \sigma^2(X)$  and

$\text{cor}[X, \ln Y] = \text{cov}[X, \ln Y] / \sigma_X \sigma_{\ln Y}$  . In the multiple regression case, the corresponding expressions are more complicated but the functional dependence remains.

<sup>21</sup> All wage and minimum wage figures below are expressed in 1979 dollars.

<sup>22</sup> In this case, the adjustment is achieved by asking, "How would the 1988 distribution of wages have been different if the minimum wage had been raised to its (real) 1979 level rather than being at its actual (real) 1988 level?" Other adjustments are made by asking similar counterfactuals, i.e., "How would the 1988 distribution of wages have been different if the variable in question had been distributed as it was in 1979 rather than as it actually was in 1988?"

<sup>23</sup> In 1979, the most important variable was gender, followed by schooling and experience with approximately equal importance. The same point holds: the relative contributions of different statistically significant variables could not have been seen from the regression equation alone.

## REFERENCES

- Ahuja, V., Bidawi, B., Ferreira, F., & Walton, M. (1997). *Everyone's Miracle?* Washington: World Bank.
- Almeida dos Reis, J. G., & de Barros, R. E. (1991). Wage Inequality and the Distribution of Education. *Journal of Development Economics*, 36(1), 117-143.
- Andersen, L. E. (2000). *Social Mobility in Latin America*. Unpublished Working Paper, Universidad Católica Boliviana.
- Arcos, X. R. (1996). *Descomposición de la Desigualdad del Consumo en Ecuador*. Masters Thesis, Cornell University, Ithaca, New York.
- Atkinson, A. B. (1997). Bringing Income Distribution in From the Cold. *Economic Journal*, 107, 297-321.
- Becker, G. S. (1964). *Human Capital*. New York: National Bureau of Economic Research.
- Becker, G. S. (1967). *Human Capital and the Personal Distribution of Income: An Analytical Approach*. W.S. Woytinski Lecture, University of Michigan.
- Behrman, J. R., Knight, J. B., & Sabot, R. H. (1983). A Simulation Alternative to the Comparative R<sup>2</sup> Approach to Decomposing Inequality. *Oxford Bulletin of Economics and Statistics*, 45, 307-312.
- Blau, F. D., & Ferber M. A. (1998). *The Economics of Women, Men, and Work*. Third Edition Englewood Cliffs, NJ: Prentice-Hall.
- Blau, F. D., & Kahn, L. M. (1997). Swimming Upstream: Trends in the Gender Wage Differential in the 1980s. *Journal of Labor Economics*, 15(1), 1-42.
- Blinder, A. S. (1973). Wage Discrimination: Reduced Form and Structural Estimates. *Journal of Human Resources*, 8, 436-455.
- Bourguignon, F., Fournier, M., & Gurgand, M. (1998). *Distribution, Development and Education: Taiwan, 1979-1992*. DELTA, processed.
- Bourguignon, F., & Martinez, M. (1997). *Decomposition of the Change in the Distribution of Primary Family Incomes: A Microsimulation Approach Applied to France, 1979-1989*. DELTA, processed.

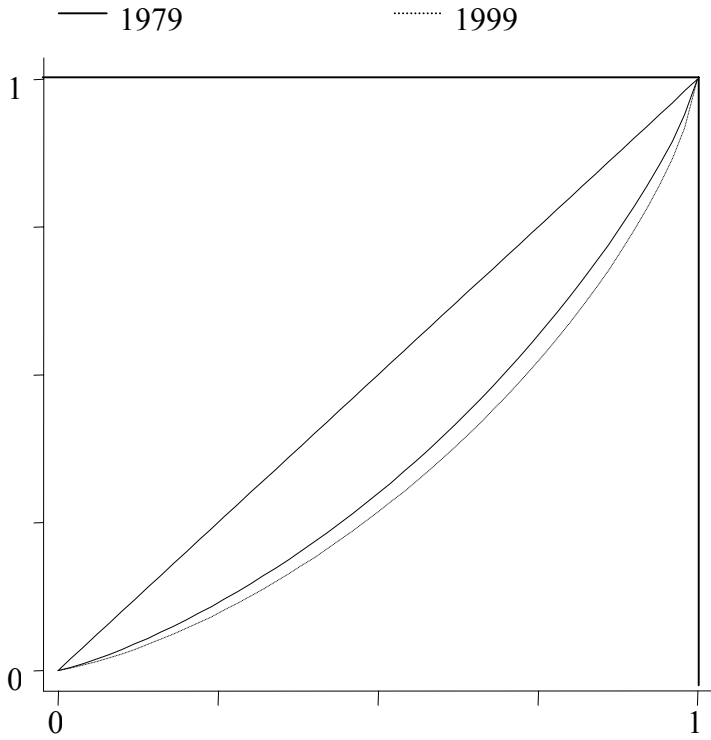
- Burt, O.R., & Finley, R. M. (1968). Statistical Analysis of Identities in Random Variables. *American Journal of Agricultural Economics*, 50, 734-744.
- Card, D. (1996). The Effect of Unions on the Structure of Wages: A Longitudinal Analysis. *Econometrica*, 64(4), 957-979.
- Chiswick, B. R., & Mincer, J. (1972). Time-Series Changes in Personal Income Inequality in the United States from 1939, with Projections to 1985. *Journal of Political Economy*, 80(3) (May/June, Part II), S34-S66.
- Contreras, D. (2000). *Explaining Wage Inequality in Chile: Does Education Really Matter?* Department of Economics, Universidad de Chile, Chile.
- Cowell, F. A., & Jenkins, S. P. (1995). How Much Inequality Can We Explain?: A Methodology and an Application to the United States. *The Economic Journal*, 105(429), 421-430.
- DiNardo, J., Fortin, N. M., & Lemieux, T. (1996). Labor Market Institutions and the Distribution of Wages, 1973-1992: A Semiparametric Approach. *Econometrica*, 64(5), 1001-1044.
- Economic Report of the President 1999* (1999). Washington: U.S. Government Printing Office.
- Fei, J. C. H., Ranis, G., & Kuo, S. W. Y. (1978). Growth and the Family Distribution of Income by Factor Components. *Quarterly Journal of Economics*, 92(1), 17-53.
- Fields, G. S., Leary, J., López Calva, L., & Pérez de Rada, E. (1998). *Education's Crucial Role in Explaining Labor Income Inequality in Urban Bolivia*. Cornell University, processed.
- Fields, G. S., & Mitchell, J. (1999). Changing Income Inequality in Taiwan: A Decomposition Analysis. In T. N. Srinivasan and G. Saxonhouse (Eds.), *Development, Duality, and the International Regime: Essays in Honor of Gustav Ranis*. Ann Arbor: University of Michigan Press.
- Fields, G. S., & Yoo, G. (2000). Falling Labor Income Inequality in Korea's Economic Growth: Patterns and Underlying Causes. *Review of Income and Wealth*, 46(2), 139-159.
- Forster, M. F., & Pellizzari, M. (2000). *Trends and Driving Factors in Income Distribution and Poverty in the OECD Area*. OECD Labour Market and Social Policy Occasional Paper No. 42.
- Foster, J. E., & Ok, E. A. (1999). Lorenz Dominance and the Variance of Logarithms. *Econometrica*, 67(4), 901-907.

- Freeman, R. (1980). Union Wage Practices and the Dispersion of Wages. *Industrial and Labor Relations Review*, 36(1), 3-21.
- Freeman, R. (1993). How Much Has De-Unionization Contributed to the Rise in Male Earnings Inequality? In S. Danziger & P. Gottschalk (Eds.), *Uneven Tides: Rising Inequality in America*. New York: Russell Sage Foundation.
- Gindling, T. H., & Trejos, J. D. (2001). *Cambios en la Desigualdad del Ingreso Laboral en Costa Rica, 1976-1999: Medidas y Causas*. Universidad de Costa Rica, Costa Rica.
- Goldberger, A. (1964). *Econometric Theory*. New York: Wiley.
- Goldberger, A. (1970). On the Statistical Analysis of Identities: Comment. *American Journal of Agricultural Economics*, 52(1), 154-155.
- Heltberg, R. (2001). *Analyzing Inequality Using Income Regressions: Vietnam, 1992-97*. Unpublished Working Paper, University of Copenhagen, Denmark.
- Juhn, C., Murphy, K. M., & Pierce, B. (1991). Accounting for the Slowdown in Black-White Wage Convergence. In Marvin Koster (Ed.), *Workers and Their Wages*. Washington, DC: American Enterprise Institute Press.
- Juhn, C., Murphy, K. M., & Pierce, B. (1993). Wage Inequality and the Rise in Returns to Skill. *Journal of Political Economy*, 101(3), 410-442.
- Katz, L. F., & Autor, D. H. (1999). Changes in the Wage Structure and Earnings Inequality. In O. Ashenfelter & D. Card (Eds.), *Handbook of Labor Economics*. Amsterdam: North-Holland.
- Katz, L. F., & Murphy, K. M. (1992). Changes in Relative Wages, 1963-1987: Supply and Demand Factors. *Quarterly Journal of Economics*, 107(1), 35-78.
- Lam, D. (1999). *Generating Extreme Inequality: Schooling, Earnings, and Intergenerational Transmission of Human Capital in South Africa and Brazil*. University of Michigan, processed.
- Lam, D., & Levison, D. (1991). Declining Inequality in Schooling in Brazil and Its Effects on Inequality in Earnings. *Journal of Development Economics*, 37, 199-225.
- Layard, R., & Zabalza, A. (1979). Family Income Distribution: Explanation and Policy Evaluation. *Journal of Political Economy*, 87(5), S133-S161.
- Levy, F. (1999). *The New Dollars and Dreams*. New York: Russell Sage Foundation.

- Mincer, J. (1958). Investment in Human Capital and Personal Income Distribution. *Journal of Political Economy*, 66, 281-302.
- Mincer, J. (1970). The Distribution of Labor Incomes: A Survey with Special Reference to the Human Capital Approach. *Journal of Economic Literature*, 8(1), 1-26.
- Mincer, J. (1974). *Schooling, Experience, and Earnings*. New York: National Bureau of Economic Research.
- Mincer, J. (1997). Changes in Wage Inequality, 1970-1990. In S. W. Polachek (Ed.), *Research in Labor Economics*, Vol. 16. Greenwich, CT: JAI Press.
- Moffitt, R. A., & Gottschalk, P. *Trends in the Autocovariance Structure of Earnings in the U.S.: 1969-1987*. Brown University and Boston College.
- Mood, A. M., Graybill, F. A., & Boes, D. C. *Introduction to the Theory of Statistics*. Third Edition, McGraw-Hill.
- Morduch, J., & Sicular, T. (1998). *Rethinking Inequality Decomposition, with Evidence from Rural China*. Harvard Institute for International Development Working Paper No. 636.
- Murphy, K. M., & Welch, F. (1992). The Structure of Wages. *Quarterly Journal of Economics*, 107(1), 285-326.
- Oaxaca, R. L. (1973). Male-Female Wage Differentials in Urban Labor Markets. *International Economic Review*, 14(3), 693-709.
- Oaxaca, R. L., & Ransom, M. R. (1994). On Discrimination and the Decomposition of Wage Differentials. *Journal of Econometrics*, 61(1), 5-21.
- Plotnick, R. D. (1982). Trends in Male Earnings Inequality. *Southern Economic Journal*, 48(3), 724-732.
- Pyatt, G., Chen, C., & Fei, J. (1980). The Distribution of Income by Factor Components. *Quarterly Journal of Economics*, 95(3), 451-473.
- Ravallion, M., & Chen, S. (1999). When Economic Reform is Faster than Statistical Reform: Measuring and Explaining Income Inequality in Rural China. *Oxford Bulletin of Economics and Statistics*, 61(1), 33-56.
- Redmund, G., & Kattuman, P. (2001). *Employment Polarisation and Inequality in the UK and Hungary*. University of New South Wales and Cambridge University, U.K.

- Robbins, D., & Gindling, T. H. (1999). Trade Liberalization and the Relative Wages for More-Skilled Workers in Costa Rica. *Review of Development Economics*, 3(2), 140-154.
- Sánchez Torrez, F., & Méndez, J. N. (1998). Descomposición de la Desigualdad del Ingreso Laboral Urbano: 1976-1997. In F. S. Torrez (Ed.), *La Distribución del Ingreso en Colombia*. Bogotá: Departamento Nacional de Planeación.
- Shorrocks, A. F. (1982). Inequality Decomposition by Factor Components. *Econometrica*, 50(1), 193-211.
- Shorrocks, A. F. (1983). The Impact of Income Components on the Distribution of Family Incomes. *Quarterly Journal of Economics*, 98(2), 311-326.
- Shorrocks, A. F. (1999). *Decomposition Procedures for Distributional Analysis: A Unified Framework Based on the Shapley Value*. University of Essex, processed.
- Welch, F. (1999). In Defense of Inequality. *American Economic Review*, 89(2), 1-17.
- Yun, M. (2002). *Earnings Inequality in USA, 1961-1999: Comparing Inequality Using Earnings Equations*. University of Western Ontario, processed.

Figure 1.  
Lorenz Curves for the Distribution of Earnings in the United States, 1979 and 1999



**Table 1.**  
**United States: Descriptive Statistics.**

Variable Group	Independent Variable	1979		1999	
		Mean	Std Dev	Mean	Std Dev
<b>Gender</b>	Male	0.62	0.48	0.56	0.50
	Female	0.38	0.48	0.44	0.50
<b>Race</b>	White	0.89	0.32	0.83	0.37
	Nonwhite	0.11	0.32	0.17	0.37
<b>Experience</b>	Potential experience	16.64	12.71	19.03	10.97
	Potential experience squared	438.48	524.31	482.59	461.27
<b>Schooling</b>	Less than complete high school	0.17	0.38	0.09	0.29
	High school grad	0.41	0.49	0.32	0.47
	Some college	0.21	0.41	0.29	0.46
	College grad and beyond	0.21	0.41	0.30	0.46
<b>Occupation</b>	Executive, professional, and technical	0.29	0.46	0.36	0.48
	Sales & admin support	0.24	0.42	0.25	0.43
	All other	0.47	0.50	0.39	0.49
<b>Industry</b>	Public administration, professional, & related services	0.28	0.45	0.31	0.46
	All other services, finance, trade, transport	0.35	0.48	0.42	0.49
	Manufacturing, construction, mining, agriculture	0.38	0.48	0.27	0.45
<b>Region</b>	Northeast	0.22	0.42	0.17	0.38
	North Central	0.27	0.44	0.24	0.43
	South	0.22	0.42	0.25	0.43
	West	0.28	0.45	0.34	0.47

**Table 2.**  
**United States: Earnings Equation Results, 1979 and 1999.**  
**Dependent Variable: Logarithm of Labor Earnings.**  
**Schooling Measured in Categories**  
(t-statistics in parentheses)

<b>Variable Group</b>	<b>Independent Variable</b>	<b>1979</b>	<b>1999</b>
<b>Gender</b>	Male	0.432	0.316
		(98.08)	(55.79)
<b>Race</b>	White	0.097	0.066
		(15.83)	(9.50)
<b>Experience</b>	Potential experience	0.033	0.040
		(62.24)	(48.50)
	Potential experience squared	-0.001	-0.001
		(-46.86)	(-34.01)
<b>Schooling</b> (Less than complete high school omitted)	High school grad	0.191	0.272
		(32.43)	(27.33)
	Some college	0.289	0.417
		(41.11)	(40.09)
	College grad and beyond	0.447	0.729
		(55.73)	(63.62)
<b>Occupation</b> (Executive, prof, and technical omitted)	Sales & admin support	-0.171	-0.233
		(-28.56)	(-30.98)
	All other	-0.226	-0.341
		(-39.03)	(-45.34)
<b>Industry</b> (Public administration, professional & related services omitted)	All other services, finance, trade, transport	0.038	0.067
		(7.34)	(10.14)
	Manufacturing, construction, mining, agriculture	0.149	0.172
		(27.64)	(23.02)
<b>Region</b> (Northeast omitted)	North Central	0.042	-0.054
		(7.78)	(-6.63)
	South	-0.080	-0.106
		(-13.91)	(-13.07)
	West	0.012	-0.085
		(2.28)	(-11.14)
Constant		5.540	5.439
		(482.59)	(330.59)
	Adjusted R <sup>2</sup>	.4146	.3833
	F-statistic	2128.02	1578.61
	N	42,045	35,579

**Table 3.**  
**United States: The Contribution of Each Explanatory Factor to**  
**Earnings Inequality and to the Change in Inequality,**  
**1979-1999.**  
(Schooling Measured in Categories)

	<b>Factor Inequality Weight of That Factor in That Year</b>		<b>Contribution of that Factor to the Change in Inequality as Measured by:</b>	
	<b><math>s_j(\ln Y)</math>, <u>1979</u></b>	<b><math>s_j(\ln Y)</math>, <u>1999</u></b>	<b><math>\pi_j</math> (Gini), <u>1979-1999</u></b>	<b><math>\pi_j</math> (log- variance), <u>1979-1999</u></b>
	(1)	(2)	(3)	(4)
Gender	0.180	0.057	-0.55	-0.22
Race	0.008	0.004	-0.02	-0.01
Experience	0.072	0.066	0.04	0.05
Schooling in categories	0.080	0.161	0.56	0.34
Occupation	0.053	0.091	0.28	0.18
Industry	0.012	0.000	-0.06	-0.03
Region	0.010	0.005	-0.02	0.00
Residual	0.585	0.617	0.77	0.69

Notes: For definition of  $s_j(\ln Y)$ , see equation (8.a).  
For definition of  $\pi_j$  (.), see equation (17.b).

**Table 4.**  
**United States: Decomposing the Contribution of Years of Education to**  
**Changing Inequality of Labor Earnings,**  
**1979 and 1999.**

	Components of Education's Factor Inequality Weight		Percentage of Change in Education's Factor Inequality Weight Explained by:		
	<u>1979</u>	<u>1999</u>	<u>(19)</u>	<u>(21)</u>	<u>(26)</u>
Education's factor inequality weight	0.089	0.161			
Coefficient on years of education	0.054	0.088	.84	1.69	1.05
Standard deviation of years of education	2.812	2.570	-.15	-.31	-.05
Correlation between labor earnings and years of education	0.301	0.438	.64	na	
Standard deviation of labor earnings	0.512	0.617	<u>-.32</u>	<u>-.64</u>	
Total			1.01	.74	1.00

**Table 5.**  
**Analyzing Rising Earnings Inequality for Women and Men Separately.**

*Part A. United States: Rising Earnings Inequality for Women, 1979-1999.*

<u>Inequality Index</u>	<u>1979</u>	<u>1999</u>
Gini Coefficient	0.236	0.317
Log-Variance	0.175	0.330

**United States: The Contribution of Each Explanatory Factor to Earnings Inequality and to the Change in Inequality for Women, 1979-1999.**

	$s_j(\ln Y)$ , <u>1979</u> (1)	$s_j(\ln Y)$ , <u>1999</u> (2)	$\pi_j$ (Gini), <u>1979-1999</u> (3)	$\pi_j$ (log-variance), <u>1979-1999</u> (4)
Race	0.000	0.000	0.00	0.00
Experience	0.028	0.041	0.08	0.06
Schooling in years	0.112	0.183	0.39	0.26
Occupation	0.131	0.137	0.15	0.14
Industry	0.000	-0.008	-0.03	-0.02
Region	0.009	0.006	0.00	0.00
Residual	0.720	0.641	0.41	0.55

**United States: Decomposing the Contribution of Years of Education to Changing Inequality of Labor Earnings for Women, 1979 and 1999.**

	<u>Components of Education's Factor Inequality Weight</u>		<u>Share of Change in Education's Factor Inequality Weight Explained by:</u>		
	<u>1979</u>	<u>1999</u>	<u>(19)</u>	<u>(21)</u>	<u>(26)</u>
Education's Factor Inequality Weight:	0.112	0.183			
Coefficient on years of schooling	0.048	0.090	1.27	2.54	.97
Standard deviation of years of schooling	2.536	2.427	-0.09	-.018	.03
Corr between earnings and schooling	0.386	0.482	0.46	na	
Standard deviation of labor earnings	0.419	0.575	<u>-0.65</u>	<u>-1.31</u>	
Total			0.98	1.05	1.000

Notes: For definition of  $s_j(\ln Y)$ , see equation (8.a).  
For definition of  $\pi_j$  (.), see equation (17.b).

*Part B. United States: Rising Earnings Inequality for Men, 1979-1999.*

<u>Inequality Index</u>	<u>1979</u>	<u>1999</u>
Gini Coefficient	0.256	0.336
Log-Variance	0.233	0.385

**United States: The Contribution of Each Explanatory Factor to Earnings Inequality and to the Change in Inequality for Men, 1979-1999.**

	$s_j(\ln Y)$ , <u>1979</u> (1)	$s_j(\ln Y)$ , <u>1999</u> (2)	$\pi_j$ (Gini), <u>1979-1999</u> (3)	$\pi_j$ (log-variance), <u>1979-1999</u> (4)
Race	0.018	0.007	-0.03	-0.01
Experience	0.125	0.093	-0.01	0.04
Schooling in years	0.115	0.166	0.33	0.24
Occupation	0.047	0.093	0.24	0.16
Industry	-0.001	-0.007	-0.03	-0.02
Region	0.013	0.005	-0.02	-0.01
Residual	0.683	0.642	0.51	0.58

**United States: Decomposing the Contribution of Years of Education to Changing Inequality of Labor Earnings for Men, 1979 and 1999.**

	<u>Components of Education's Factor Inequality Weight</u>		<u>Share of Change in Education's Factor Inequality Weight Explained by:</u>		
	<u>1979</u>	<u>1999</u>	<u>(19)</u>	<u>(21)</u>	<u>(26)</u>
Education's Factor Inequality Weight:	0.115	0.166			
Coefficient on years of schooling	0.057	0.086	1.10	2.20	1.05
Standard deviation of years of schooling	2.966	2.671	-0.29	-0.57	-.05
Corr between earnings and schooling	0.326	0.448	0.87	na	
Standard deviation of labor earnings	0.483	0.621	<u>-0.68</u>	<u>-1.37</u>	
Total			.99	.25	1.00

Notes: For definition of  $s_j(\ln Y)$ , see equation (8.a).  
For definition of  $\pi_j$  (.), see equation (17.b).

**Appendix Table 1.**  
**United States: Earnings Equation Results, 1979 and 1999.**  
**Dependent Variable: Logarithm of Labor Earnings.**  
**Schooling Measured in Categories**  
(t-statistics in parentheses)

Variable Group	Independent Variable	1979	1999
<b>Gender</b>	Male	0.432	0.316
		(98.96)	(55.9)
<b>Race</b>	White	0.098	0.071
		(15.94)	(10.13)
<b>Experience</b>	Potential experience	0.032	0.039
		(60.84)	(47.14)
		Potential experience squared	-0.001
		(-44.89)	(-32.75)
<b>Schooling</b>	High school grad	0.054	0.088
		(60.23)	(70.65)
<b>Occupation</b> (Executive, prof, and technical omitted)	Sales & admin support	-0.164	-0.236
		(-28.14)	(-31.92)
		All other	-0.222
		(-39.46)	(-47.72)
<b>Industry</b> (Public administration, professional & related services omitted)	All other services, finance, trade, transport	0.046	0.079
		(8.82)	(11.80)
		Manufacturing, construction, mining, agriculture	0.157
		(29.19)	(24.70)
<b>Region</b> (Northeast omitted)	North Central	0.043	-0.055
		(7.83)	(-6.78)
		South	-0.077
		(-13.44)	(-13.37)
	West	0.021	-0.074
		(3.81)	(-9.65)
Constant		5.080	4.678
		(304.02)	(199.50)
	Adjusted R <sup>2</sup>	0.4200	0.3842
	F-statistic	2538.46	1850.63
	N	42,045	35,579

**Appendix Table 2.**  
**United States: The Contribution of Each Explanatory Factor to**  
**Earnings Inequality and to the Change in Inequality,**  
**1979-1999.**  
(Schooling Measured in Years)

	<b>Factor Inequality Weight of That Factor in That Year</b>		<b>Contribution of that Factor to the Change in Inequality as Measured by:</b>	
	$s_j(\ln Y),$ <u>1979</u>	$s_j(\ln Y),$ <u>1999</u>	$\pi_j$ (Gini), <u>1979-1999</u>	$\pi_j$ (log- variance), <u>1979-1999</u>
	(1)	(2)	(3)	(4)
Gender	0.180	0.057	-0.55	-0.22
Race	0.008	0.004	-0.02	-0.01
Experience	0.070	0.064	0.03	0.05
Schooling in categories	0.089	0.161	0.52	0.32
Occupation	0.051	0.094	0.30	0.19
Industry	0.012	0.000	-0.06	-0.03
Region	0.010	0.005	-0.02	0.00
Residual	0.580	0.616	0.79	0.70

Notes: For definition of  $s_j(\ln Y)$ , see equation (8.a).  
For definition of  $\pi_j$  (.), see equation (17.b).

