Lorenz Decomposition and Interdistributional Lorenz Comparisons*

JOHN A. BISHOP
DEPARTMENT OF ECONOMICS
EAST CAROLINA UNIVERSITY

K. VICTOR CHOW
DEPARTMENT OF FINANCE
WEST VIRGINIA UNIVERSITY

LESTER A. ZEAGER
DEPARTMENT OF ECONOMICS
EAST CAROLINA UNIVERSITY

September 25, 2003

* The authors wish to thank John Formby, Peter Lambert, and Jacques Silber for helpful comments and suggestions.

Lorenz Decomposition and Interdistributional Lorenz Comparisons

— Abstract —

We use recently developed methods to perform decompositions of the Lorenz curve in the United States by race, region, and marital status. The decomposed Lorenz ordinates are used to construct interdistributional Lorenz curves (ILCs), which allow us to identify an economic advantage by one subgroup over another or changes in economic advantage over time. We propose asymptotically distribution-free estimators for the ILCs and apply these estimators to data from the Current Population Survey for 1977 and 1997. As one might expect, there are economic advantages by race, region, and marital status, even in 1997. Economic advantage is greatest for marital status and smallest for region in both years. We find significant convergence (i.e., a smaller economic advantage) over time by race and region, though not by marital status.

1. INTRODUCTION

Researchers examining income inequality by population subgroup can choose from a wide array of decomposable inequality indices. The results from this approach may depend on the index selected and there is no basis on which to select a particular index (Conrad 1999).[1] To bypass the problem of selecting an index, we propose a dominance approach to subpopulation decomposition. Following Bishop, Chow, and Zeager (2003), we decompose the Lorenz curve itself, the most general measure of inequality (Atkinson 1970). From these decomposed Lorenz ordinates, we construct the interdistributional Lorenz curves (ILCs) proposed by Butler and McDonald (1987). We also derive estimators for the standard errors of the interdistributional Lorenz ordinates and provide tests for stochastic dominance relations among ILCs.

ILCs are created by plotting the corresponding incomplete moments for two population subgroups at common income targets. The moments used here correspond to proportions of subgroup *income* below given income targets. If the moments are equal at every income target, there is no interdistributional inequality (i.e., the plot corresponds to the 45-degree line). If the moments for the groups are not equal at every income target, such that one group is unambiguously disadvantaged, and we plot the moments for the disadvantaged group on the horizontal axis, the ILC will lie below the 45-degree line.

As Deutsch and Silber (1999) point out, an ILC can capture an "economic advantage" by one population subgroup over another. In this paper ILCs measure the relative concentration of incomes among the poor in different population subgroups. In our empirical analysis, we compare economic advantage across subgroups in the United States defined by race, region, and marital status. Moreover, we check for convergence (smaller economic advantage) or divergence (larger economic advantage) over time in ILCs for each category. All comparisons are made using micro data from the Current Population Survey (CPS) for 1977 and 1997. We look only at primary families, excluding single-person families and unrelated individuals.

We find economic advantages for families in which the head is white, lives outside the South, and is married. In both years the advantages are greatest for marital status and smallest for region. We also find convergence (i.e., dominance of 1977 by 1997) for region and race, but not for marital status.

The rest of the paper is organized as follows. Section 2 describes the methods we use for generating ILCs and the statistical inference procedures we use for the dominance tests. Section 3 illustrates the approach using CPS data from the United States. Section 3 offers brief concluding remarks.

2. METHODS

The idea of comparing inequality *between* distributions (rather than income dispersion *within* one distribution) goes back to Gini (1916, 1959), who developed the notions of "Transvariazione" (the extent of overlapping) and "Ipertransvariazione" (the difference between the concentrations of each distribution below a reference point in the other distribution). Subsequent work by Dagum (1980, 1987), Shorrocks (1982), Vinod (1985), Gastwirth (1985), and Deutsch and Silber (1997, 1999) extended and refined the ideas. Deutsch and Silber (1999) provide additional references and a useful survey of the literature. Butler and McDonald (1987) proposed the first dominance approach, though even they introduced a summary (Pietra) index in their empirical analysis. This paper breaks new ground by implementing a dominance approach exclusively.[2]

We begin with the concepts that provide the foundation for ILCs. Let x be a continuous income variable with a probability density $f(x)$ and let $F(x)$ denote the cumulative distribution function (CDF) of x. Let the inverse CDF of x be written as $0 \leq F^{-1}(p) \leq \infty$ and without loss of generality, let $\tau = F^{-1}(p)$ define target incomes. If p = 0.1, 0.2, …, 1.0, then the target incomes become the decile order statistics. Let the indicator variable $I_\tau^x = 1$ if $x \leq \tau$ and $I_\tau^x = 0$ otherwise.

Given a target income τ, we can define the *h*-th partial moment for $x < \tau$ of the density function $f(x)$ as

(1) $$M(\tau, h, x) = \int_0^\tau x^h f(x)dx = \int_0^\infty (xI_\tau^x)^h dF(x) = E\left[(xI_\tau^x)^h\right],$$

where E is the expectation operator. For $h = 0$ the partial moment reduces to $F(\tau)$, which involves accumulating population shares. Following Butler and McDonald (1987), we define the normalized incomplete moment of x for $x \leq \tau$ as

$$\phi(\tau, h, x) = M(\tau, h, x) / E(x^h), \tag{2}$$

where $E(x^h) = \lim_{\tau \to \infty} M(\tau, h, x)$. For $h = 1$ the normalized incomplete moment is a Lorenz ordinate, which becomes clear if we write the Lorenz curve in the form proposed by Bishop, Chow, and Formby (1994),

$$\phi(\tau, 1, x) = \mu_x^{-1} \int_0^\tau x f(x) dx = \mu_x^{-1} \int_0^\infty x I_\tau^x dF(x) = E\left[x I_\tau^x\right] / E(x), \tag{3}$$

where $E(x)$ is the mean of x. Thus, we interpret $\phi(\tau, 1, x)$ as the proportion of income in x received by individuals with incomes x less than or equal to a target income τ.

Butler and McDonald (1987) use $\phi(\tau, h, x)$ to define ILCs for population subgroups. They plot $\phi_\alpha(\tau, h, x) = \phi(\tau, h, x^{(\alpha)})$ against $\phi_\beta(\tau, h, x) = \phi(\tau, h, x^{(\beta)})$ for either h = 0 or h = 1, where α and β are subgroups of interest within the population. If $\phi_\alpha(\tau, h, x) = \phi_\beta(\tau, h, x)$ at each τ, then the ILC corresponds to the 45-degree line. When one subgroup (α) is unambiguously disadvantaged [$\phi_\alpha(\tau, h, x) > \phi_\beta(\tau, h, x)$ at each τ], the ILC lies below the 45-degree line. The interpretation of an ILC depends on the parameter h. As Butler and McDonald (1987, 14) note, the ILC comparison for h = 0 is equivalent to a first-order dominance comparison. That is, the ILC for h = 0 lies below the 45-degree line if

$$\varphi_\beta(\tau, 0, x) = F_\beta(\tau) < F_\alpha(\tau) = \phi_\alpha(\tau, 0, x) \tag{4}$$

for all τ. For h = 1, the ILC lies below the 45-degree line if

(5) $$\phi_\beta(\tau,1,x) = S_\beta(\tau) < S_\alpha(\tau) = \phi_\alpha(\tau,1,x),$$

where

(6) $$S_i(\tau,x) = \mu_i^{-1}\int_0^\tau x f_i(x)dx, \text{ for i = α, β}$$

is the share of income in subgroup i going to income units in subgroup i with incomes not exceeding τ.[3] In this case the ILC captures the relative concentration of income among the poor for the two subgroups.

To illustrate these ideas, consider a numerical example. Suppose we have the following vector of incomes for a population, where the underlined incomes belong to Beta (β) and other incomes belong to Alpha (α):

(7) $$x = [\ 1, 1, \underline{2}, 3, 4, \underline{4}, 5, \underline{10}, \underline{10}, 15, \underline{20}, \underline{25}\ ].$$

For convenience, we choose target incomes that correspond to the population quartiles, $(\tau_1 = 2, \tau_2 = 4, \tau_3 = 10, \tau_4 = 25)$. Table 1 reports the values for the Lorenz ordinates and their decomposed components [i.e., $\phi(\tau,1,x)$, $\phi_\alpha(\tau,1,x)$, and $\phi_\beta(\tau,1,x)$] at each of the pre-selected target incomes.

By choosing target incomes corresponding to quartiles of the population, it is possible to interpret the values of $\phi(\tau,1,x)$ in Table 1 as ordinates of the familiar Lorenz curve, i.e., the bottom 25 percent of the population $(x \le x_1 = 2)$ receives 4 percent of the total income. Similar interpretations can be given to normalized incomplete moments for subgroups Alpha and Beta, $\phi_\alpha(\tau,1,x)$ and $\phi_\beta(\tau,1,x)$. That is, 6.9 percent of total Alpha income and 2.8 percent of total Beta income goes to persons at or below the \$2 target income. Bishop, Chow, and Zeager (2002) point out that $\phi_\alpha(\tau,1,x)$ and $\phi_\beta(\tau,1,x)$, weighted by their income shares, sum to $\phi(\tau,1,x)$.

From Table 1, we create an ILC by plotting $\phi_\beta(\tau,1,x)$ against $\phi_\alpha(\tau,1,x)$ as in Figure 1. As $\phi_\alpha(\tau,1,x) > \phi_\beta(\tau,1,x)$ at each target income τ, the ILC lies below the 45-degree line. Following Deutsch and Silber (1987), we infer that Beta has an economic advantage over Alpha, because Alpha's incomes are more concentrated in the lower income classes (at every target income) than Beta's incomes.

To express these ideas more generally, let incomes be classified by *K* mutually exclusive groups $\{\Phi_k, k = 1,2,..,K\}$ and define an indicator variable G_k^x such that $G_k^x = 1$ if $x \in \Phi_k$ and $G_k^x = 0$ otherwise. This indicator variable allows us to rewrite (3), because $E(xI_\tau^x | G_k^x = 1) = E(xG_k^x I_\tau^x)/E(G_k^x)$ and $E(x | G_k^x = 1) = E(xG_k^x)/E(G_k^x)$. Bishop, Chow, and Zeager (2002) use this approach to show that:

THEOREM 1. *$\phi(\tau,1,x)$ can be decomposed by $\phi(\tau,1,x^{(k)})$ for k = 1, 2, …, K in that*

(8)
$$\phi(\tau,1,x) = \sum_{k=1}^{K} P^{(k)} \cdot \phi(\tau,1,x^{(k)}),$$

where $P^{(k)} = E\left[x \cdot G_k^x\right] / E(x)$. We can interpret $P^{(k)}$ as the income share of subgroup k with respect to the income variable *x*.

To develop an inference test for ILCs, we select a set of m income classes or target income levels, denoted by $\{\tau_i | i = 1,2,...,m\}$, to which there correspond *K* sets of decomposed Lorenz ordinates $\{\phi(\tau_i,1,x^{(k)}) \mid i = 1, 2, ..., m, \text{ and } k = 1,2,...,K\}$.[4] Let

($x_1, x_2, \ldots, x_N$) be a set of identical and independently distributed (i.i.d.) random sample incomes drawn from the population density *f(x)*. By equation (3) and Theorem 1, the decomposed Lorenz ordinates (DLOs) can be estimated as

$$(9) \qquad \hat{\phi}_{i,(k)} = \hat{\phi}(\tau_i, 1, x^{(k)}) = \left[N^{-1} \sum_{j=1}^{N} x_j G_k^{x_j} I_{\tau_i}^{x_j} \right] \Big/ \left[N^{-1} \sum_{j=1}^{N} x_j G_k^{x_j} \right]$$

Let $\underset{1x(mK)}{\Phi} = (\Phi_1, \Phi_2, ..., \Phi_{mK})' = \left((\phi_{1,(1)}, \ldots, \phi_{m,(1)}), (\phi_{1,(2)}, \ldots, \phi_{m,(2)}), \ldots, (\phi_{1,(K)}, \ldots, \phi_{m,(K)}) \right)'$ be a vector of mK DLOs. Let the estimates of the vector Φ can be written as $\underset{1x(mK)}{\hat{\Phi}} = (\hat{\Phi}_1, \hat{\Phi}_2, ..., \hat{\Phi}_{mK})' = \left((\hat{\phi}_{1,(1)}, \ldots, \hat{\phi}_{m,(1)}), (\hat{\phi}_{1,(2)}, \ldots, \hat{\phi}_{m,(2)}), \ldots, (\hat{\phi}_{1,(K)}, \ldots, \hat{\phi}_{m,(K)}) \right)'$. By equations (3) and (5), the DLOs are functions of $E\left[xG_k^x I_{\tau_i}^x \right]$ and $E\left[xG_k^x \right]$ for *i* = 1,2, …, *m* and *k* = 1,2,…,*K*. To derive the asymptotic sampling distribution of $\overline{\Phi}$, we must determine the sampling distributions of these estimates, $\overline{xG_k^x I_{\tau_i}^x}$, and $\overline{xG_k^x}$ for *i* = 1,2, …, *m* and *k* = 1,2,…,*K*.

We define the vector of $K(m+1)$ parameter estimators as

$\underset{1x[K(m+1)]}{\overline{\Psi}} = (\overline{\psi}_1, \overline{\psi}_2, ..., \overline{\psi}_{K(m+1)})' = \left(\left(\overline{xG_1^x I_{\tau_1}^x}, \ldots, \overline{xG_1^x I_{\tau_m}^x}, \overline{xG_1^x} \right), \ldots, \left(\overline{xG_K^x I_{\tau_1}^x}, \ldots, \overline{xG_K^x I_{\tau_m}^x}, \overline{xG_K^x} \right) \right)'$.

THEOREM 2. *Suppose* $(x_1, x_2, \ldots, x_N)$ *are i.i.d. random samples of a size of N drawn from the population density function f(x). Given a set of predefined target incomes* $\{\tau_i | i = 1,2,3,\ldots,m\}$ *such that* $0 < \tau_1 < \ldots < \tau_m < \infty$, *and a population decomposed into K mutually exclusive groups, the vector* $\sqrt{N}(\overline{\Psi} - \Psi)$ *converges in distribution to a K(m+1) variate normal distribution with mean zero and a variance-covariance* $\Omega = (\sigma_{i,j})$, *where*

$$\underset{(K(m+1))x(K(m+1))}{\Omega} = \begin{bmatrix} [\alpha_{ij}^{11}] & \cdots & [\alpha_{ij}^{1K}] \\ \vdots & \ddots & \vdots \\ [\alpha_{ij}^{K1}] & \cdots & [\alpha_{ij}^{KK}] \end{bmatrix}, \textit{ and}$$

$$[\alpha_{ij}^{kl}]_{(m+1)x(m+1)} = \begin{cases} Cov[(xG_k^x I_{\tau_i}^x), (xG_l^x I_{\tau_j}^x)] & \text{for } i, j \le m \\ Cov[(xG_k^x I_{\tau_i}^x), (xG_l^x)] & \text{for } i \le m, \quad j = (m+1) \\ Cov[(xG_k^x), (xG_l^x I_{\tau_j}^x)] & \text{for } i = (m+1), \ j \le m \\ Cov[(xG_k^x), (xG_l^x)] & \text{for } i = j = (m+1) \end{cases},$$

and where *Cov* denotes the covariance.

PROOF. Given that income samples *x* and the indicator variables of *G* and *I* are *i.i.d.*, the *h*-th power function *i.i.d* random variable is also *i.i.d.*. From direct calculations, it can be shown that $E(\overline{\psi}_i) = \psi_i$, $i = 1,2,\ldots, K(m+1)$. Then, for large samples, the Kolmogorov Strong Law of Large Numbers implies that $\overline{\psi}_i$ converges in probability to ψ_i. By the Lindeberg-Levy Central Limit Theorem, it follows that $\sqrt{N}(\overline{\psi}_i - \psi_i)$ converges in distribution to $N(0, \sigma_i^2)$. Finally, by the Cramer-Wald Theorem, it can be shown that $\sqrt{N}(\overline{\Psi} - \Psi)$ converges to a multivariate normal distribution, $N(\underline{0}, \Omega)$. Q.E.D.

Theorem 2 allows us to analyze the sampling distribution of the DLOs. Then, applying Rao's (1965) theorem on the limiting distribution of differentiable functions of random variables, the limiting distribution of $\hat{\Phi}$ is also multivariate normal. We summarize this result in the following theorem.

THEOREM 3. *Under the conditions of Theorem 3, the vector of DLOs* $\underset{1x(mK)}{\hat{\Phi}} = (\hat{\Phi}_1,\ \hat{\Phi}_2, ..., \hat{\Phi}_{mK})' = \left((\hat{\phi}_{1,(1)}, \ldots, \hat{\phi}_{m,(1)}), (\hat{\phi}_{1,(2)}, \ldots, \hat{\phi}_{m,(2)}), \ldots, (\hat{\phi}_{1,(K)}, \ldots, \hat{\phi}_{m,(K)}) \right)'$ *is asymptotically normal in that* $\sqrt{N}(\hat{\Phi} - \Phi)$ *has a limiting* Km *-variate normal distribution with mean zero and covariance matrix* $V = J\Omega J' = (v_{ij})$*, where* Ω *is defined in Theorem 2 and J is defined as* $\underset{(2Km)x(2Km+k)}{J} = \left[\delta\overline{\Phi_j} / \delta\overline{\Psi_j} \right]_{\overline{\Psi}=\Psi}$.

Then, the covariance estimate of the *k*-th and *l*-th DLOs, $\hat{\phi}_{j,(k)}$ and $\hat{\phi}_{j,(l)}$ can then be determined as follows:

$$
\begin{aligned}
\overline{Cov}(\hat{\phi}_{j,(k)}, \hat{\phi}_{j,(l)}) = a_k a_l & \left[\overline{(xG_k^x I_{\tau_j}^x)\ (xG_l^x I_{\tau_j}^x)} - \left(\overline{xG_k^x I_{\tau_j}^x}\right)\left(\overline{xG_l^x I_{\tau_j}^x}\right) \right] + \\
c_{jk} c_{jl} & \left[\overline{(xG_k^x)(xG_l^x)} - \left(\overline{xG_k^x}\right)\left(\overline{xG_l^x}\right) \right] + \\
a_k c_{jk} & \left[\overline{(xG_k^x I_{\tau_j}^x)(xG_k^x)} - \left(\overline{xG_k^x I_{\tau_j}^x}\right)\left(\overline{xG_k^x}\right) \right] + \\
a_l c_{jl} & \left[\overline{(xG_l^x I_{\tau_j}^x)(xG_l^x)} - \left(\overline{xG_l^x I_{\tau_j}^x}\right)\left(\overline{xG_l^x}\right) \right]
\end{aligned}
$$

where $a_k = \left(\overline{xG_k^x}\right)^{-1}$, $a_l = \left(\overline{xG_l^x}\right)^{-1}$, $c_{jk} = -\left(\overline{xG_k^x I_{\tau_j}^x}\right)\left(\overline{xG_k^x}\right)^{-2}$, and $c_{jl} = -\left(\overline{xG_l^x I_{\tau_j}^x)}\right)\left(\overline{(xG_l^x)}\right)^{-2}$ for $j = 1, 2, ..., m$.

Theorem 3 provides the full variance-covariance structure of the asymptotic normal distribution of a vector of DLOs. We see that it depends only on the first and second moments and can be consistently estimated without any prior specification of the population density underlying the sample data. Thus, statistical inference for hypothesis testing of ILCs ─ which are created from DLOs ─ is distribution-free and statistically straightforward.

For example, if we wanted to test whether the ILC for two subgroups (say, 1 and 3) is significantly different from a 45-degree line, the null hypothesis could be defined as

$H_0^1: D = \underline{0}$, where $\hat{D} = (\hat{\phi}_{1,(1)} - \hat{\phi}_{1,(3)}, \ldots, \hat{\phi}_{m,(1)} - \hat{\phi}_{m,(3)})'$ is a $1xm$ vector of sample estimates of differences in the DLOs for the relevant subgroups. From Theorem 3, the test statistic

$$\chi^2 = N\hat{D}' I \hat{V}^{-1} I' \hat{D}, \qquad (10)$$

where

$$I = \begin{bmatrix} \begin{bmatrix} 1 & 0 & -1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & -1 & \cdots & 0 \end{bmatrix} \end{bmatrix}_{m \text{ x } Km}$$

and $\hat{V}$ is the estimate of the *(Km)x(Km)* full variance-covariance matrix in Theorem 3, is asymptotically distributed as a (central) chi-squared variate with *m*-degrees of freedom.

Alternatively, a multiple-comparison test could be performed by defining the null hypothesis as $H_0^2: \hat{D}_j = (\hat{\phi}_{j,(1)} - \hat{\phi}_{j,(3)}) = 0$ for $j = 1,2,\ldots, m$. By calculating a set of standard normal test statistics, $Z_j = \hat{D}_j / S_j$, the standard errors S_j can be obtained easily from Theorem 3. The asymptotic joint confidence interval of at least 100(1-α) percent for the set of $\hat{D}_j$, $j = 1,2,\ldots, m$, is (under H_0^2): $Z_j \pm SMM(\alpha; m; \infty)$ for $j=1,2,\ldots,m$, where $SMM(\alpha; m; \infty)$ is the asymptotic critical value of the α-point of the *studentized maximum modulus* (*SMM*) distribution with parameter m and ∞ degrees of freedom.[5] This approach allows us to perform dominance tests and is the statistical inference method featured in the paper.

3. DATA

The data for our empirical analysis are drawn from the 1977 and 1997 Current Population Survey (CPS). Selected sample statistics are shown in Table 2. Clearly, the mean and standard deviation of family income increased substantially between 1977 and 1997. However, since the ILC is based solely on income proportions of the subgroups in the overall income distribution for a given year, our preliminary results are not corrected for inflation. We also make no adjustments for the top-coding of incomes or for the size and composition of the family, but our sample is restricted to primary families (i.e., it excludes single-person families and unrelated individuals).

Table 2 also reveals a reduction in the proportion of family heads that were married, along with a smaller reduction in the proportion of whites and a slight increase in the proportion living in the South. In addition to a change in overall sample size, these shifts would lead to differences in the sample sizes of the population subgroups by race, region, and marital status over time.

4. APPLICATIONS

This section estimates ILCs for three socioeconomic categories that have long been associated with inequality in the United States: race, region, and marital status. The ILCs can be used to measure economic advantage *within* the three categories and to make comparisons *across* categories. Also, comparisons of ILCs over time allow us to detect widening or narrowing of economic advantage within a single category. This section illustrates comparisons of each kind.

We first use the ILCs and associated inference tests to measure economic advantages between households with heads who are white and non-white, married and non-married, and from the South and non-South. Butler and McDonald (1987) use black-white income comparisons to motivate the original presentation of ILCs. The economic disadvantages of single parent households are of interest to many researchers. Bishop, Formby and Thistle (1992) discuss the extensive literature on economic differences between the South and non-South.

The analysis begins in Table 3a, which presents the South versus non-South comparisons. Columns 1 and 2 in Table 3a report income shares below the population deciles for 1977. In column 1, 1.41 percent of family income for the non-South is below the target income for the first decile of the overall U.S. population, while 2.30 percent of family income for the South is below this target income in Column 2. Further inspection shows that for every decile in 1977 the income shares for the non-South are smaller than those for the South. Figure 2a plots column 1 against column 2 from Table 3a to obtain the South versus non-South ILC for 1977. Likewise, Figure 2a plots column 4 against column 5 from Table 3a to show the South versus non-South ILC for 1997.[6]

Having constructed Figure 2a, it is natural to inquire whether each ILC differs significantly from the 45-degree line. Columns 3 and 6 in Table 3a report the multiple-comparison tests based on H_0^2. The tests compute the differences between the subgroup income shares below target incomes, using the null hypothesis of no economic advantage. Column 3 gives the differences and their standard errors in 1977, with stars (*) indicating significant differences at the five-percent level. There are stars at each decile in 1977, so we reject the null hypothesis of no economic advantage and conclude that the 1977 ILC lies below the 45-degree line. Column 6 gives the corresponding information for 1997 and leads to a similar conclusion.[7]

Figure 2a indicates that the South versus non-South ILC may be converging toward the 45-degree line over time. Table 3a offers an informal convergence test for ILCs in column 7.[8] The null hypothesis for this test is that the ILC for 1997 lies above the ILC for 1977 at every decile. Rejection of the null hypothesis indicates that there is either no change in economic advantage over time or a crossing of the ILCs. A crossing ILC, like the familiar Lorenz curve crossing, precludes an unambiguous conclusion. The convergence test compares column 3 to column 6. If the difference between the ILC and 45-degree line diminishes significantly over time, we can infer that the ILC is converging to interdistributional equality.

Column 7 of Table 3a reveals dominance of 1977 by 1997 for the South versus the Non-South. In other words, the Non-South's economic advantage over the South diminished significantly at the bottom five deciles of the income distribution for the overall population during this period. Note that one could not easily anticipate this pattern from inspection of Figure 2a.

Table 3b provides the equivalent information for the white versus non-white comparisons. The corresponding ILCs are shown in Figure 2b. In both years we find that the ILCs are significantly different from the 45-degree line (see columns 3 and 6 in Table 3b). Column 7 in Table 3b provides the statistical test for white versus non-white convergence. At each decile except the highest two, the test statistic is significant at the five percent level, implying dominance of 1977 by 1997.

Table 3c presents the remaining comparison: households with married and non-married heads. Not surprisingly, the households with married heads have an economic advantage over those with non-married heads in both years. Figure 2c makes this point clearly, but it also suggests the possibility of *divergence* over time. Column 7 of Table 3c reports both positive and negative signs and some of each are statistically significant. We can infer from this pattern that the 1977 and 1997 ILCs cross, so the test for convergence is inconclusive. That is, at the bottom of the distribution the 1997 ILC is nearer the 45-degree line, while at the top of the distribution the 1977 ILC is nearer.

Figure 3a shows the ILCs for racial, regional, and marital categories in 1977. Observe that the ILC for marital status lies below the ILC for race, which indicates that the economic advantage of the married over the non-married is greater than the economic advantage of whites over non-whites. Also, Figure 3 reveals that the ILC for region lies above the ILC for race, which indicates that the economic advantage of the Non-South over the South is smaller than the economic advantage associated with race or marital status. Figure 3b shows the patterns for 1997, which are similar to those for 1977.

Table 4 reports the statistical tests for dominance *across* socioeconomic categories. We find less interdistributional inequality by region than by race or marital status ─ and by race than by marital status ─ in 1977. The same results continue to hold in 1997, so the patterns are stable over time.

5. CONCLUSIONS

We decompose Lorenz curves for the United States in 1977 and 1997 by race, region, and marital status. We then use the decomposed Lorenz ordinates to construct interdistributional Lorenz curves (ILCs). The ILCs are compared in 1997 across racial, regional, and marital categories to evaluate the relative economic advantages across the groups. We find that economic advantage is greatest by marital status and smallest by region. We also compare ILCs over time *within* each socioeconomic category. These comparisons show convergence for race and region, but an ambiguous outcome for marital status.

REFERENCES

Bishop, J. A., K. V. Chow, and J. P. Formby, (1994) "Testing for Marginal Changes in Income Distributions with Lorenz and Concentration Curves," *International Economic Review*, 35, 479-488.

Bishop, J. A., K. V. Chow, and L. A. Zeager, (2003) "Decomposing Lorenz and Concentration Curves," *International Economic Review*, 44, 965-978.

Bishop, J. A., K. V. Chow, and L. A. Zeager, (2003) "Changes in Economic Advantage by National Origin After German Reunification," *Estadistica* (Journal of the Inter-American Statistical Institute), forthcoming.

Bishop, J. A., J. P. Formby, and P. D. Thistle, (1992) "The Convergence of South and Non-South Income Distributions, 1969-1979," *American Economic Review*, 82, 262-272.

Butler, R. J., and J. B. McDonald, (1987) "Interdistributional Income Inequality," *Journal of Business and Economic Statistics*, 5, 13-18.

Conrad, C., (1999) "Inequality Decomposition by Population Subgroup and the Analysis of Interdistributional Inequality: Comment," in J. Silber, ed., *Handbook of Income Inequality Measurement*, (Dortrecht: Kluwer), 397-403.

Dagum, C. (1980) "Inequality Measures Between Income Distributions with Applications," *Econometrica*, 48, 1791-1803.

Dagum, C. (1987) "Measuring the Economic Affluence Between Populations of Income Receivers," *Journal of Business and Economic Statistics*, 5, 5-12.

Deutsch, J., and J. Silber, (1997) "Gini's 'Transvariazione' and the Measurement of Inequality: Within and Between Distributions," *Empirical Economics*, 22, 547-554.

Deutsch, J., and J. Silber, (1999) "Inequality Decomposition by Population Subgroup and the Analysis of Interdistributional Inequality", in J. Silber (Ed.), *Handbook of Income Inequality Measurement* (Dordrecht: Kluwer) (pp. 363-397).

Gastwirth, J. L. (1985) "Comment on 'Measurement of Economic Distance Between Blacks and Whites' by H.D. Vinod," *Journal of Business and Economic Statistics*, 3, 405-409.

Gini, C. (1916) Il Concetto di "Transvariazione" e le sue prime applicazioni," Studi di Economia, Finanza e Statistica, editi del Giornali degli Economisti e Revista de Statistica, Reprinted in Gini, 1959.

Gini, C. (1959) *Memorie de Metadologia Statistica: Volume Secondo – Transvariazione.* Rome: Libreria Goliardica.

Lambert, P. (1993) "Inequality Reduction through the Income Tax," *Economica,* 60, 357-65.

.

Rao, C. R., (1965) *Linear Statistical Inference and its Application*, (New York: Wiley).

Shorrocks, A. F. (1982) "On the Distance Between Income Distributions," *Econometrica*, 50, 1337-1339.

Shorrocks, A. F. (1999) "Decomposition Procedures for Distributional Analysis: A Unified Framework Based on the Shapley Value," Unpublished Manuscript, University of Essex and Institute for Fiscal Studies.

Stoline, M. R., and H. K. Ury, (1979) "Tables of the Studentized Maximum Modulus Distribution and an Application to Multiple Comparisons Among Means, *Technometrics*, 21, 87-93.

Vinod, H. D. (1985) "Measurement of Economic Distance Between Blacks and Whites," *Journal of Business and Economic Statistics*, 3, 78-88.

ENDNOTES

[1] See Deutsch and Silber (1999) for a discussion of alternative decomposable indices. For an interesting attempt to provide a unified framework for the decomposition of indices, see Shorrocks (1999)

[2] Bishop, Chow and Zeager (2003) use ILCs to construct Gini summary measures of economic advantage but do not apply inference testing.

[3] Lambert (1993, 359) states a similar (necessary) condition, called (a*) in his paper, for an income tax to achieve a reduction in overall inequality.

[4] This approach allows us to relax the assumption about the continuous CDF, since the Lorenz and concentration ordinates correspond to a set of target incomes instead of a set of quantile functions. Empirically, the targets are selected as a set of sample quantiles ($\hat{\xi}_p$) of the income variable x, i.e., $p_1 = 0.1$, $p_2 = 0.2$, …, $p_9 = 0.9$, which are the sample deciles. Let a random sample of size N be given from the population. If the CDF of x is strictly monotonic, then $\hat{\xi}_p$ has the property of strong or almost sure consistency (Rao 1965, 335).

[5] The *SMM* table can be found in Stoline and Ury (1979).

[6] We use the scatterplot procedure in RATS to generate the ILCs.

[7] We also performed Chi-squared tests for each case in tables 3a-3b. the results are consistent with the multiple comparison tests.

[8] A more formal test of ILC convergence to the 45-degree line would require us to create a pooled distribution of incomes across time, which would require us to make corrections for inflation. We hope to provide such a test in the future.

Table 1
Normalized Incomplete Moments at Target Incomes

Target Income[a] x	Normalized Incomplete Moment: Population $\phi(\tau,1,x)$	Normalized Incomplete Moment: Alpha $\phi_{\alpha}(\tau,1,x)$	Normalized Incomplete Moment: Beta $\phi_{\beta}(\tau,1,x)$
2	0.04	0.069	0.028
4	0.15	0.310	0.085
10	0.40	0.483	0.366
25	1.00	1.000	1.000

[a] The target incomes are set at population quartiles. Thus, $\phi(\tau,1,x)$ gives the familiar Lorenz ordinates for the population.

Table 2
Sample Statistics[a]

Variable	1977	1997
Family Income[b]	18,035.52 (12,151.35)	56,061.97 (54,607.08)
Proportion of Whites	0.89	0.86
Proportion Living in South	0.30	0.31
Proportion Married	0.84	0.77
Number of Observations	41,011	34,900

[a] Numbers in parentheses are standard deviations.
[b] Incomes are not adjusted for top-coding or differences in family size or composition.

Table 3a
Interdistributional Lorenz Ordinates (Income Shares) for South vs. Non-South: 1977 and 1997

Population Decile	1977			1997			Convergence Statistic[a]
	Non-South (1)	South (2)	Difference (3)=(2)-(1)	Non-South (4)	South (5)	Difference (6)=(5)-(4)	(3) vs. (6)
1	.0141	.0230	.0090* (.0008)	.0117	.0157	.0040* (.0006)	5.00
2	.0473	.0675	.0201* (.0017)	.0391	.0506	.0115* (.0015)	3.79
3	.0953	.1274	.0321* (.0027)	.0790	.1000	.0210* (.0024)	3.07
4	.1583	.2025	.0443* (.0037)	.1325	.1624	.0299* (.0036)	2.79
5	.2372	.2893	.0521* (.0047)	.2011	.2360	.0350* (.0047)	2.57
6	.3318	.3890	.0572* (.0057)	.2829	.3304	.0476* (.0061)	1.15
7	.4447	.5002	.0555* (.0065)	.3864	.4332	.0468* (.0074)	0.88
8	.5788	.6281	.0493* (.0071)	.5149	.5543	.0394* (.0086)	0.89
9	.7447	.7751	.0304* (.0071)	.6820	.7026	.0206* (.0096)	0.82

[a]The convergence statistic is the Student Maximum Modulus (SMM). The five-percent critical value for the SMM statistic is 2.77.

Table 3b
Interdistributional Lorenz Ordinates (Income Shares) for White vs. Non-White: 1977 and 1997

Population Decile	1977			1997			Convergence Statistic[a] (3) vs. (6)
	White (1)	Non-White (2)	Difference (3)=(2)-(1)	White (4)	Non-White (5)	Difference (6)=(5)-(4)	
1	.0132	.0561	.0429* (.0021)	.0108	.0303	.0195 (.0013)	9.47
2	.0462	.1311	.0849* (.0040)	.0380	.0799	.0419 (.0027)	8.91
3	.0946	.2173	.1227* (.0057)	.0790	.1372	.0582 (.0042)	9.10
4	.1588	.3070	.1482* (.0073)	.1329	.2116	.0786 (.0059)	7.42
5	.2391	.3986	.1595* (.0087)	.2019	.2911	.0892 (.0075)	6.12
6	.3350	.4945	.1595* (.0098)	.2857	.3909	.1052 (.0092)	4.03
7	.4474	.6076	.1602* (.0107)	.3883	.5009	.1127 (.0107)	3.14
8	.5815	.7200	.1385* (.0109)	.5147	.6253	.1105 (.0121)	1.74
9	.7446	.8528	.1082* (.0100)	.6775	.7762	.0987 (.0127)	0.59

[a]The convergence statistic is the Student Maximum Modulus (SMM). The five-percent critical value for the SMM statistic is 2.77.

Table 3c
Interdistributional Lorenz Ordinates (Income Shares) for Married vs. Non-Married: 1977 and 1997

Population Decile	1977			1997			Convergence Statistic[a]
	Married (1)	Non-Married (2)	Difference (3)=(2)-(1)	Married (4)	Non-Married (5)	Difference (6)=(5)-(4)	(3) vs. (6)
1	.0094	.0808	.0714* (.0023)	.0061	.0594	.0533* (.0017)	6.40
2	.0380	.1872	.1493* (.0042)	.0270	.1480	.1210* (.0034)	5.39
3	.0825	.2992	.2167* (.0059)	.0607	.2529	.1922* (.0051)	3.14
4	.1436	.4122	.2686* (.0073)	.1084	.3662	.2578* (.0066)	1.12
5	.2216	.5211	.2995* (.0083)	.1731	.4729	.2998* (.0079)	-0.03
6	.3173	.6199	.3026* (.0090)	.2551	.5829	.3278* (.0089)	-1.98
7	.4316	.7154	.2838* (0093)	.3580	.6889	.3309* (.0096)	-3.52
8	.5693	.7999	.2305* (.0092)	.4876	.7929	.3053* (.0099)	-5.53
9	.7380	.8888	.1508* (.0084)	.6597	.8820	.2223* (.0097)	-5.57

[a]The convergence statistic is the Student Maximum Modulus (SMM). The five-percent critical value for the SMM statistic is 2.77.

Table 4: Comparing Interdistributional Lorenz Curves Across Groups[a]

	South	White	Married
South	*	S (1997)	S (1997)
White	S (1977)	*	W (1997)
Married	S (1977)	W (1977)	*

[a] "S" implies less interdistributional income inequality in South vs. Non-South comparison than in the comparisons for other groups. "W" implies less interdistributional income inequality in White vs. Non-White comparison than in the comparisons for other groups. All the tests are significant at the 5% percent level.

Figure 1
Interdistributional Lorenz Curve

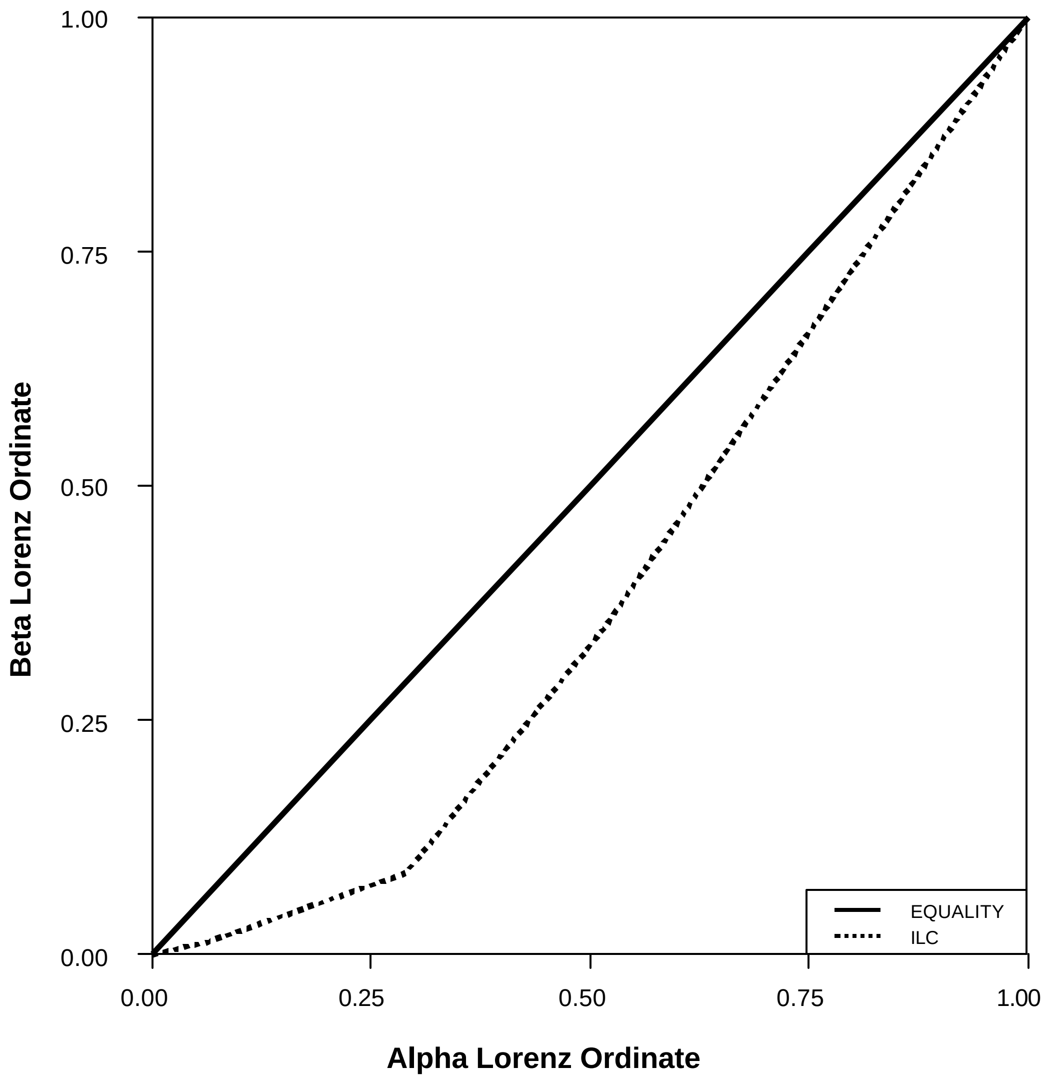

Figure 2a
Interdistributional Lorenz Curves (Income Shares) for the South and NonSouth: 1977 and 1997

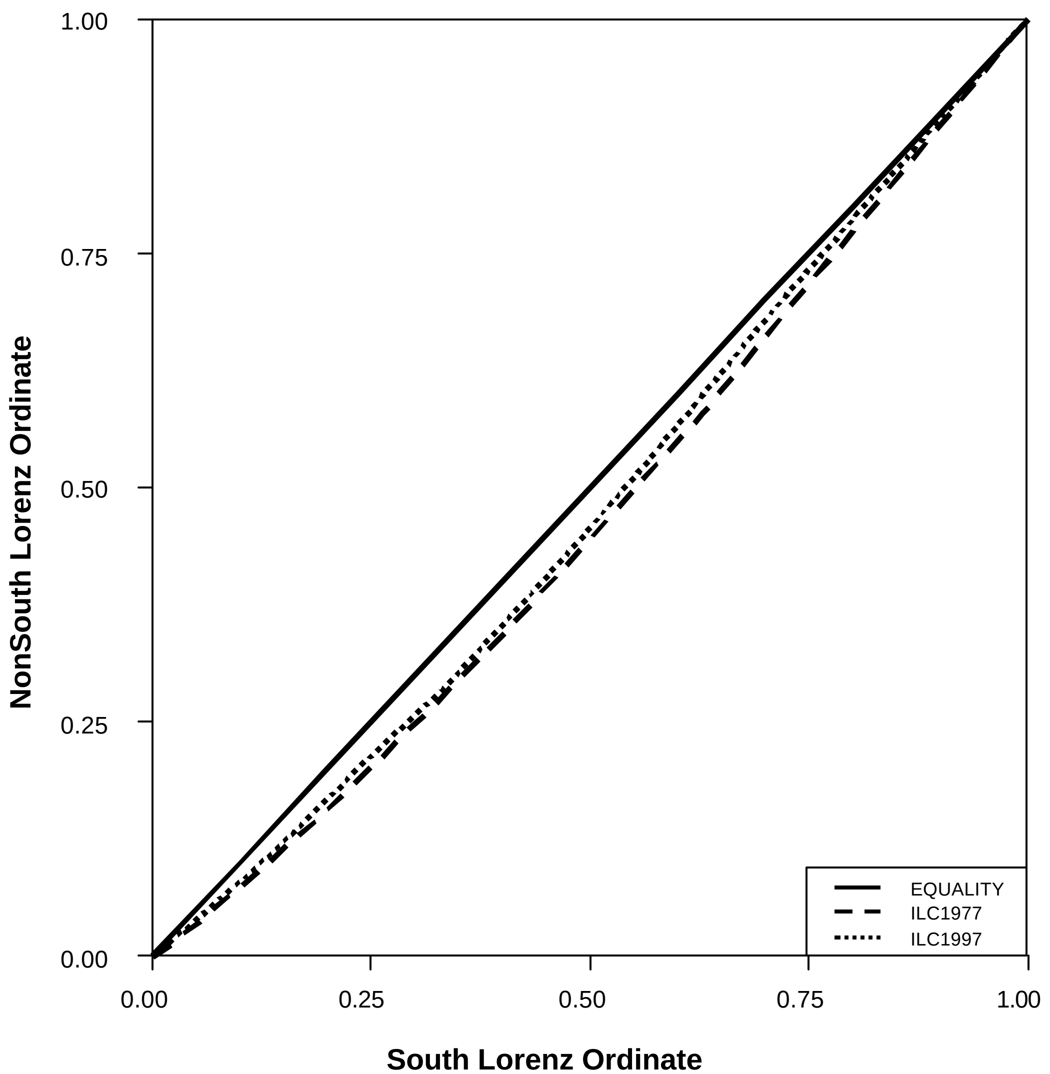

Figure 2b
Interdistributional Lorenz Curves (Income Shares) for the White and NonWhite: 1977 and 1997

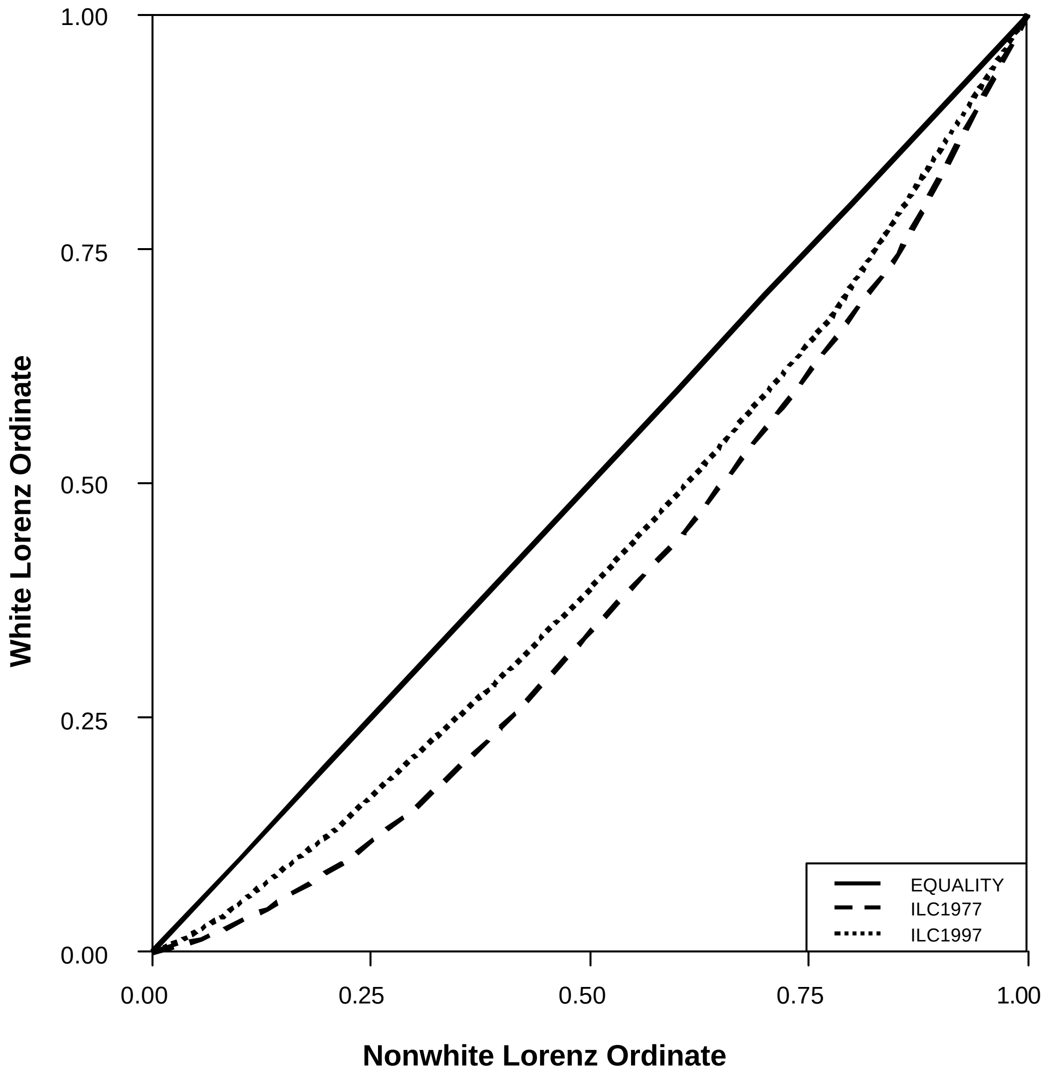

Figure 2c
Interdistributional Lorenz Curves (Income Shares) for the Married and NonMarried: 1977 and 1997

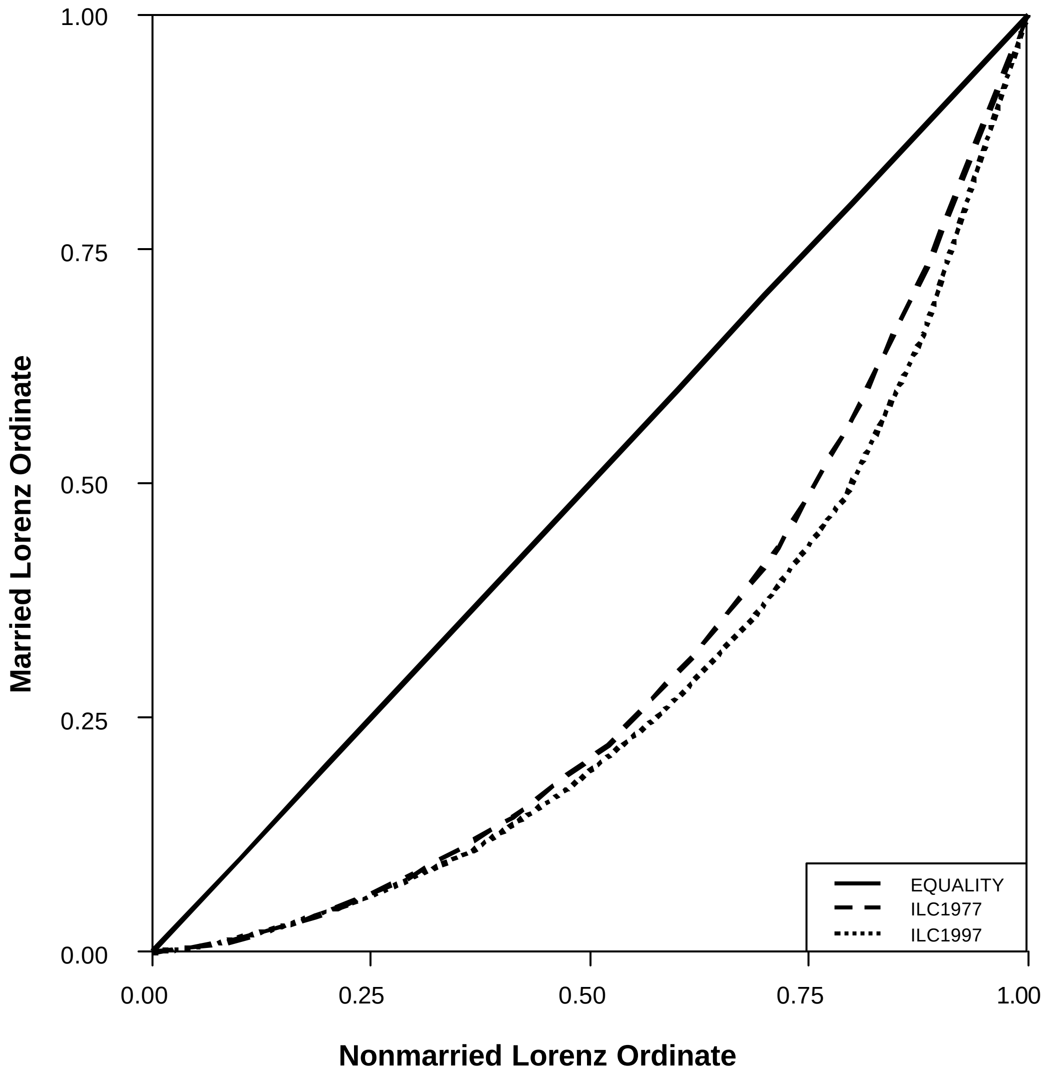

Figure 3a
Interdistributional Lorenz Curves (Income Shares) by Race, Region, and Marital Status in 1977

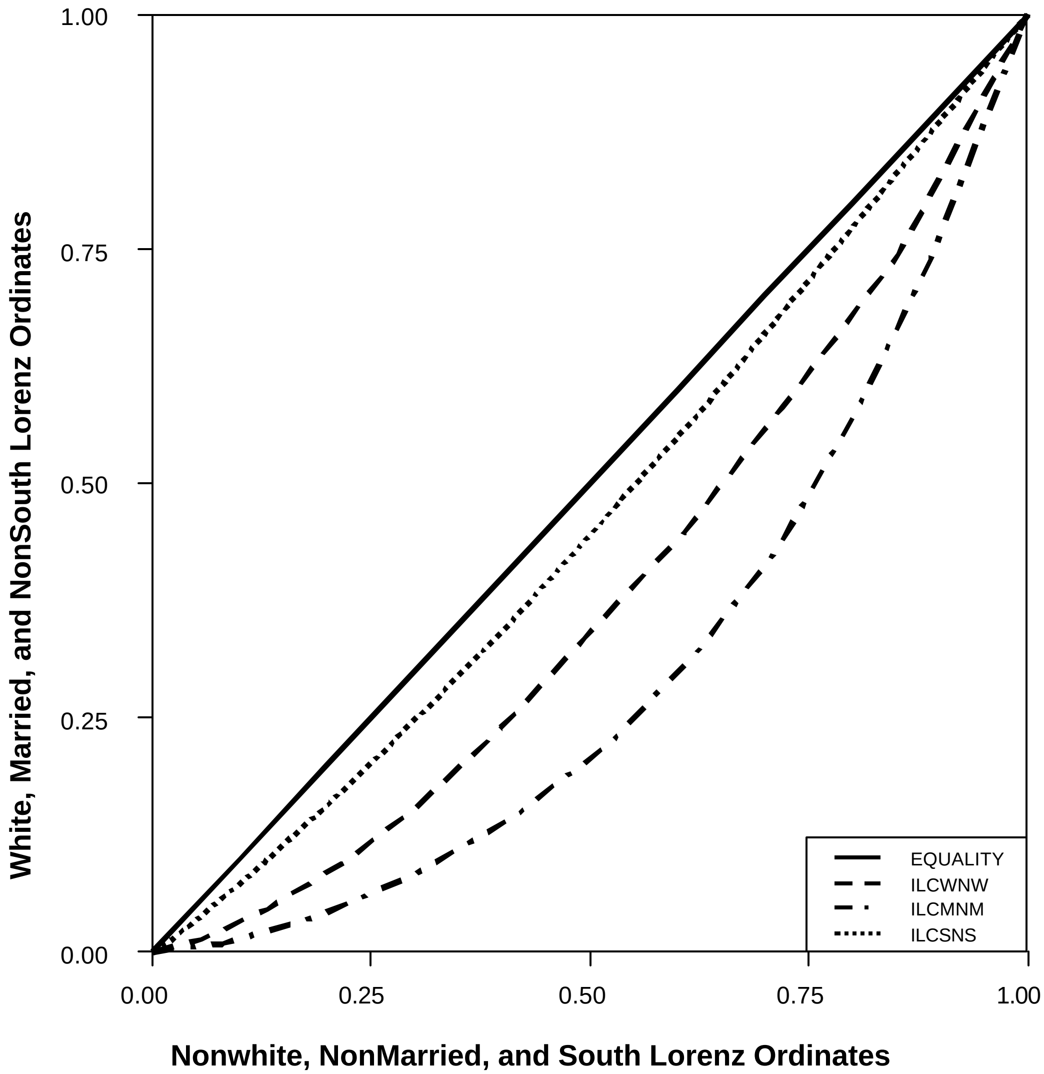

Figure 3b
Interdistributional Lorenz Curves (Income Shares) by Race, Region, and Marital Status in 1997