

# Distribution-Free Statistical Inference for Generalized Lorenz Dominance Based on Grouped Data

Y. Murasawa<sup>a</sup> and K. Morimune<sup>b</sup>

<sup>a</sup>College of Economics, Osaka Prefecture University, Sakai, Osaka 599-8531, Japan  
(murasawa@eco.osakafu-u.ac.jp)

<sup>b</sup>Institute of Economic Research, Kyoto University, Kyoto 606-8501, Japan  
(morimune@kier.kyoto-u.ac.jp)

**Abstract:** One income distribution is preferable to another under any increasing and Schur-concave (S-concave) social welfare function if and only if the generalized Lorenz (GL) curve of the first distribution lies above that of the second. Thus testing for GL dominance of one distribution over another is of interest. The paper focuses on inference based on grouped data and makes two contributions: (i) it gives a new formula for the asymptotic variance-covariance matrix of a vector of sample GL curve ordinates, interpreting it as a method-of-moments estimator, and (ii) it proposes a new test for multivariate inequality restrictions, of which GL dominance is a special case. The testing problem is  $H_0 : \theta_0 \geq 0$  vs.  $H_1 : \theta_0 \not\geq 0$ , where  $\theta_0$  is the difference between two vectors of ordinates from two GL curves, or equivalently  $H_0 : \theta_{\min} \geq 0$  vs.  $H_1 : \theta_{\min} < 0$ , where  $\theta_{\min}$  is the minimum component of  $\theta_0$ . Given the asymptotic distribution of  $\hat{\theta}_n$ , the difference between two vectors of ordinates from two sample GL curves, one can simulate the distribution of  $\hat{\theta}_{n,\min}$ , the minimum component of  $\hat{\theta}_n$ , under the least favorable case in  $H_0$  and evaluate the asymptotic  $p$ -value. For the Japanese household income data grouped into deciles, the test accepts the null hypothesis that income distribution in Japan improved from 1979 to 1994.

**Keywords:** Generalized Lorenz curve; Stochastic dominance; Method of moments; Testing multivariate inequality restrictions

## 1. INTRODUCTION

Among countless empirical works that compare income or wealth inequality across regions or over time using estimates of various inequality measures, few report the standard errors; thus we can evaluate the sampling errors in few cases. Although several procedures for statistical inference exist in the literature, few empirical researchers follow such procedures. This may be because those procedures are somewhat complicated.

Among various criteria for comparing income (or wealth) distributions, this paper focuses on the generalized Lorenz (GL) curve for two rea-

sons. First, Shorrocks [1983] shows that one income distribution is preferable to another under any increasing and Schur-concave (S-concave) social welfare function if and only if the GL curve of the first distribution lies above that of the second (GL dominance). Second, given the asymptotic distribution of a vector of sample GL curve ordinates, we can derive the asymptotic distributions of the associated vector of sample Lorenz curve ordinates and of the associated estimators of the Gini coefficient by the delta method; see Beach and Davidson [1983].

This paper makes two contributions to the literature on statistical inference for GL dominance. First, we derive the asymptotic distribution of a

vector of sample GL curve ordinates, interpreting it as a method-of-moments (MM) estimator, and obtain a new formula for its asymptotic variance-covariance matrix. Beach and Davidson [1983] apply the asymptotic theory of linear functions of order statistics to obtain a different formula. Since the asymptotic theory of MM estimators is familiar to econometricians and empirical researchers while that of linear functions of order statistics is not, our result is more intuitive. Although the MM estimator we consider is not differentiable with respect to the parameter vector, we can apply empirical process theory to derive its asymptotic distribution; see Andrews [1994].

Second, we propose a new test for multivariate inequality restrictions, of which GL dominance is a special case. Let  $\theta_0$  be a parameter vector, e.g., the difference between two vectors of ordinates from two GL curves. Consider testing

$$H_0 : \theta_0 \geq 0 \quad \text{vs.} \quad H_1 : \theta_0 \not\geq 0.$$

Let  $\theta_{\min}$  be the minimum component of  $\theta_0$ . Then we want to test

$$H_0 : \theta_{\min} \geq 0 \quad \text{vs.} \quad H_1 : \theta_{\min} < 0.$$

Let  $\hat{\theta}_n$  be a consistent and uniformly asymptotically normal estimator of  $\theta_0$ . We use  $\hat{\theta}_{n,\min}$ , the minimum component of  $\hat{\theta}_n$ , as the test statistic. Given the asymptotic distribution of  $\hat{\theta}_n$ , we can simulate the asymptotic distribution of  $\hat{\theta}_{n,\min}$  under the least favorable case in  $H_0$ , i.e.,  $\theta_0 = 0$ , and evaluate the asymptotic  $p$ -value.

An important feature of our method is that it is feasible even when only grouped data are available. As an example, we apply our method to the publicly available grouped data of the National Survey of Family Income and Expenditure in Japan. The test accepts the null hypothesis that income distribution in Japan improved from 1979 to 1994.

## 2. GENERALIZED LORENZ DOMINANCE

### 2.1 Generalized Lorenz Curves

Let  $X$  be a positive random variable. Let  $F : \mathfrak{R} \rightarrow [0, 1]$  be the cumulative distribution function (cdf) of  $X$ . Let for all  $\alpha \in [0, 1]$ ,  $x_\alpha$  be the  $100\alpha$  percentile of  $X$  defined as  $x_\alpha := \inf\{x \in \mathfrak{R}_+ : F(x) \geq \alpha\}$ . Let  $\mu := E(X)$ .

**Definition 1** *The Lorenz curve of  $X$  is  $L : [0, 1] \rightarrow [0, 1]$  such that for all  $\alpha \in [0, 1]$ ,*

$$L(\alpha) := \frac{E([X \leq x_\alpha]X)}{\mu}.$$

**Definition 2** *The generalized Lorenz (GL) curve of  $X$  is  $GL : [0, 1] \rightarrow [0, \mu]$  such that for all  $\alpha \in [0, 1]$ ,*

$$GL(\alpha) := E([X \leq x_\alpha]X).$$

Let  $F_1(\cdot)$  and  $F_2(\cdot)$  be cdfs. We say that  $F_1(\cdot)$  GL dominates  $F_2(\cdot)$  if the GL curve of  $F_1(\cdot)$  lies above that of  $F_2(\cdot)$ . For continuous random variables, GL dominance is equivalent to the second-order stochastic dominance (SSD); see Foster and Shorrocks [1988] and Yitzhaki and Olkin [1991].

### 2.2 Income Distributions and Social Welfare

#### 2.2.1 Social welfare functions

Let  $y \in \mathfrak{R}^n$  be a distribution of income (or consumption, wealth, etc.) among  $n$  households in an economy. Let  $W : \mathfrak{R}^n \rightarrow \mathfrak{R}$  be a social welfare function (SWF) that depends solely on  $y$ .

**Definition 3**  *$B \in \mathfrak{R}_+^{n \times n}$  is bistochastic if the components in each row and column add up to 1 respectively.*

**Definition 4**  *$W(\cdot)$  is Schur-concave (S-concave) if for all  $y$  and for all bistochastic matrices  $B$ ,*

$$W(By) \geq W(y).$$

S-concave functions are symmetric, i.e., for all  $y$  and for all permutation matrices  $P$ ,  $W(Py) = W(y)$ ; see Berge [1963, p. 220]. S-concave SWFs satisfy the Pigou-Dalton (P-D) principle of transfers. To be precise, an SWF is strictly S-concave if and only if it satisfies the P-D principle; see Sen [1997, p. 134]. For example, symmetric quasiconcave functions are S-concave; see Dasgupta et al. [1973, p. 183].

#### 2.2.2 Generalized Lorenz dominance and social welfare

Assume that  $y \geq 0$  and that it is ordered. The GL curve of  $y$  is for all  $\alpha \in [0, 1]$ ,

$$GL(\alpha) := \frac{1}{n} \sum_{i=1}^{[\alpha n]} y_i,$$

where  $[\cdot]$  rounds up a real number to an integer. We say that  $y$  GL dominates  $y'$  if the GL curve of  $y$  lies above that of  $y'$ .

**Theorem 1 (Shorrocks [1983])**  $W(y) \geq W(y')$  for all increasing and  $S$ -concave  $W(\cdot)$  if and only if  $y$  GL dominates  $y'$ .

Suppose that  $W(\cdot)$  is invariant to replication of the population. Then the theorem holds even when the dimensions of  $y$  and  $y'$  differ.

### 3. SAMPLE GENERALIZED LORENZ CURVES

#### 3.1 Sample Generalized Lorenz Curves

Let  $(X_1, \dots, X_n)$  be a sample of size  $n$ . Let  $\hat{F}_n : \mathfrak{R} \rightarrow [0, 1]$  be the empirical cdf given the sample, i.e., for all  $x \in \mathfrak{R}$ ,

$$\hat{F}_n(x) := \frac{1}{n} \sum_{i=1}^n [X_i \leq x].$$

Let  $X_{(1)}, \dots, X_{(n)}$  be the order statistics. Let for all  $\alpha \in [0, 1]$ ,  $\hat{x}_{n,\alpha}$  be the sample  $100\alpha$  percentile, i.e.,

$$\begin{aligned} \hat{x}_{n,\alpha} &:= \inf \left\{ x \in \mathfrak{R}_+ : \hat{F}_n(x) \geq \alpha \right\} \\ &= \inf \left\{ x \in \mathfrak{R}_+ : \frac{1}{n} \sum_{i=1}^n [X_i \leq x] \geq \alpha \right\} \\ &= X_{([\alpha n])}. \end{aligned}$$

Let  $\hat{\mu}_n$  be the sample mean.

**Definition 5** The sample GL curve given  $(X_1, \dots, X_n)$  is  $\hat{GL}_n : [0, 1] \rightarrow [0, \hat{\mu}_n]$  such that for all  $\alpha \in [0, 1]$ ,

$$\hat{GL}_n(\alpha) := \frac{1}{n} \sum_{i=1}^n [X_i \leq \hat{x}_{n,\alpha}] X_i.$$

#### 3.2 Consistency

Gail and Gastwirth [1978] prove pointwise consistency of the sample GL curves in their proof of pointwise consistency of the sample Lorenz curves.

**Theorem 2** Suppose that

- $X_1, \dots, X_n$  are independent and identically distributed (iid),
- $E(|X_1|) < \infty$ ,
- $F(\cdot)$  is strictly increasing and  $C^0$  at  $x_\alpha$ .

Then

$$\lim_{n \rightarrow \infty} \hat{GL}_n(\alpha) = GL(\alpha) \quad a.s.$$

*Proof.* See Gail and Gastwirth [1978, p. 788].

The first condition holds for simple random sampling (SRS) and probability-proportional-to-size (PPS) sampling *with* replacement. Given the third condition, which implies an infinite population, it also holds for SRS and PPS sampling *without* replacement, including systematic sampling with randomized order of the population. It does not hold for stratified sampling, however.

#### 3.3 Asymptotic Distribution

Let  $0 < \alpha_1 < \dots < \alpha_k = 1$ . Let for  $j = 1, \dots, k$ ,  $x_j$  be the  $100\alpha_j$  percentile of  $X$ . The corresponding GL curve ordinates of  $X$  are for  $j = 1, \dots, k$ ,

$$GL_j := E([X \leq x_j]X).$$

Let for  $j = 1, \dots, k$ ,  $\hat{x}_{n,j}$  be the sample  $100\alpha_j$  percentile. The corresponding sample GL curve ordinates given  $(X_1, \dots, X_n)$  are for  $j = 1, \dots, k-1$ ,

$$\hat{GL}_{n,j} := \frac{1}{n} \sum_{i=1}^n [X_i \leq \hat{x}_{n,j}] X_i,$$

and

$$\hat{GL}_{n,k} := \frac{1}{n} \sum_{i=1}^n X_i.$$

Beach and Davidson [1983] derive the asymptotic joint distribution of sample GL curve ordinates using the asymptotic theory of linear functions of order statistics, noting that for  $j = 1, \dots, k$ ,

$$\hat{GL}_{n,j} = \frac{1}{n} \sum_{i=1}^{[\alpha_j n]} X_{(i)}.$$

Since this asymptotic theory may be unfamiliar to econometricians and empirical researchers, we give an alternative derivation, noting that a vector of sample GL curve ordinates is a method-of-moments (MM) estimator.

Let

$$\theta_0 := \begin{pmatrix} x_1 \\ \vdots \\ x_{k-1} \\ GL_1 \\ \vdots \\ GL_k \end{pmatrix}, \quad \hat{\theta}_n := \begin{pmatrix} \hat{x}_{n,1} \\ \vdots \\ \hat{x}_{n,k-1} \\ \hat{GL}_{n,1} \\ \vdots \\ \hat{GL}_{n,k} \end{pmatrix}.$$

Let  $\Theta \subset \mathfrak{R}_+^{2k-1}$  be the parameter space. Given  $\theta \in \Theta$ , let for  $i = 1, \dots, n$ ,

$$m(X_i; \theta) := \begin{pmatrix} [X_i \leq x_1] - \alpha_1 \\ \vdots \\ [X_i \leq x_{k-1}] - \alpha_{k-1} \\ [X_i \leq x_1]X_i - GL_1 \\ \vdots \\ [X_i \leq x_{k-1}]X_i - GL_{k-1} \\ X_i - GL_k \end{pmatrix}.$$

Assume that  $X_1, \dots, X_n$  are iid. Let  $m_0 : \Theta \rightarrow \mathfrak{R}^{2k-1}$  be such that for all  $\theta \in \Theta$ ,

$$m_0(\theta) := E(m(X_1; \theta)).$$

Then we have a moment restriction such that

$$m_0(\theta_0) = 0. \quad (1)$$

Let  $\bar{m}_n(\cdot)$  be the sample analog of  $m_0(\cdot)$ , i.e., for all  $\theta \in \Theta$ ,

$$\bar{m}_n(\theta) := \frac{1}{n} \sum_{i=1}^n m(X_i; \theta).$$

Note that for  $j = 1, \dots, k-1$ ,

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^n [X_i \leq \hat{x}_{n,j}] &= \frac{[\alpha_j n]}{n} \\ &= \alpha_j + \frac{[\alpha_j n] - \alpha_j n}{n}. \end{aligned}$$

Hence

$$\bar{m}_n(\hat{\theta}_n) = O(n^{-1}). \quad (2)$$

Thus  $\hat{\theta}_n$  is an MM estimator of  $\theta_0$ . Theorem 2 essentially gives a sufficient condition for  $\hat{\theta}_n$  to be consistent for  $\theta_0$ .

Since  $m(\cdot; \cdot)$  is not differentiable with respect to  $\theta$ , we apply empirical process theory to derive the asymptotic distribution of  $\hat{\theta}_n$ ; see Andrews [1994]. Let  $\nu_n(\cdot)$  be a  $(2k-1) \times 1$  empirical process on  $\Theta$  given  $(X_1, \dots, X_n)$  such that for all  $\theta \in \Theta$ ,

$$\nu_n(\theta) := \frac{1}{\sqrt{n}} \sum_{i=1}^n (m(X_i; \theta) - E(m(X_i; \theta))).$$

**Theorem 3** Suppose that

- $X_1, \dots, X_n$  are iid,
- $E(|X_1|^2) < \infty$ ,
- $F(\cdot)$  is strictly increasing and  $C^1$  on its support,
- $\{\nu_n(\cdot)\}_{n=1}^\infty$  is stochastically equicontinuous.

Then

$$\sqrt{n}(\hat{\theta}_n - \theta) \rightarrow_d N(0, J^{-1}VJ^{-1}'),$$

where

$$\begin{aligned} J &:= m'_0(\theta_0), \\ V &:= \text{var}(m(X_1; \theta_0)). \end{aligned}$$

*Proof.* Available from the authors.

In our case, it turns out that the first two conditions are sufficient for stochastic equicontinuity of each component of  $\{\nu_n(\cdot)\}_{n=1}^\infty$ ; hence the last condition is unnecessary. (The details are available from the authors.)

Since the sample GL curve ordinates are the last  $k$  components of  $\hat{\theta}_n$ , it is now straightforward to obtain their asymptotic joint distribution. Let

$$GL := \begin{pmatrix} GL_1 \\ \vdots \\ GL_k \end{pmatrix}, \quad \hat{GL}_n := \begin{pmatrix} \hat{GL}_{n,1} \\ \vdots \\ \hat{GL}_{n,k} \end{pmatrix}.$$

**Theorem 4** Suppose that

- $X_1, \dots, X_n$  are iid,
- $E(|X_1|^2) < \infty$ ,
- $F(\cdot)$  is strictly increasing and  $C^1$  on its support.

Then

$$\sqrt{n}(\hat{GL}_n - GL) \rightarrow_d N(0, \Sigma),$$

where for  $i, j = 1, \dots, k$  such that  $i \leq j$ ,

$$\begin{aligned} \sigma_{i,j} &:= x_i \alpha_i (1 - \alpha_j) x_j - x_i (GL_i - \alpha_i GL_j) \\ &\quad - (GL_i - GL_i \alpha_j) x_j \\ &\quad + E([X_1 \leq x_i] X_1^2) - GL_i GL_j. \end{aligned}$$

*Proof.* Available from the authors.

The form of the asymptotic variance-covariance matrix is intuitive. The last two terms equal  $\text{cov}([X_1 \leq x_i] X_1, [X_1 \leq x_j] X_1)$ , which would have resulted if we knew the true percentiles. The first three terms capture the effect of using the sample percentiles instead of the true ones.

Compare our result with the corresponding result in Beach and Davidson [1983, Theorem 1]. In our notation, they obtain for  $i, j = 1, \dots, k$  such that  $i \leq j$ ,

$$\begin{aligned} \sigma_{i,j} &:= \alpha_i [\text{var}(X|X \leq x_i) \\ &\quad + (1 - \alpha_j)(x_i - \mu_i)(x_j - \mu_j) \\ &\quad + (x_i - \mu_i)(\mu_j - \mu_i)], \end{aligned}$$

where  $\mu_i := E(X|X \leq x_i)$ . It is tedious but straightforward to show that the two are equivalent.

### 3.4 Covariance Matrix Estimation

We can consistently estimate  $\Sigma$  by replacing the parameters associated with  $F(\cdot)$  in its expression with those associated with  $\hat{F}_n(\cdot)$ , i.e., their sample analogs. Thus we use  $\hat{\Sigma}_n$  such that for  $i, j = 1, \dots, k$  such that  $i \leq j$ ,

$$\begin{aligned} \hat{\sigma}_{n,i,j} &:= \hat{x}_{n,i} \hat{\alpha}_i (1 - \hat{\alpha}_j) \hat{x}_{n,j} \\ &\quad - \hat{x}_{n,i} \left( \hat{G}L_{n,i} - \hat{\alpha}_i \hat{G}L_{n,j} \right) \\ &\quad - \left( \hat{G}L_{n,i} - \hat{G}L_{n,i} \hat{\alpha}_j \right) \hat{x}_{n,j} \\ &\quad + \hat{E}_n \left( [X_1 \leq x_i] X_1^2 \right) - \hat{G}L_{n,i} \hat{G}L_{n,j}. \end{aligned}$$

## 4. TESTING FOR GENERALIZED LORENZ DOMINANCE

### 4.1 Multivariate One-Sided Tests and Multivariate Inequality Tests

Let  $GL_1$  and  $GL_2$  be vectors of GL curve ordinates of two distributions. Let  $\theta_0 := GL_1 - GL_2$ . Then  $\theta_0 \geq 0$  if and only if the first distribution GL dominates the second. Goldberger [1992] distinguishes the following two formulations for testing multivariate inequality hypotheses.

**Definition 6** A multivariate one-sided testing problem is

$$H_0 : \theta_0 = 0 \quad \text{vs.} \quad H_1 : \theta_0 \geq 0.$$

**Definition 7** A multivariate inequality testing problem is

$$H_0 : \theta_0 \geq 0 \quad \text{vs.} \quad H_1 : \theta_0 \not\geq 0.$$

In general, the first formulation is better for asserting  $\theta_0 \geq 0$ . A drawback of this formulation, however, is that neither hypothesis covers crossing GL curves. This is a serious drawback in our context, because it is quite possible that two GL curves cross and hence the two distributions are incomparable. Thus we choose the second formulation. Note that now we assert GL dominance by accepting the null hypothesis. Such a conclusion is weak, because the power of the test is not under our direct control.

### 4.2 Test Statistic

Let  $\hat{G}L_{1,n_1}$  and  $\hat{G}L_{2,n_2}$  be vectors of sample GL curve ordinates of two independent random samples, of sizes  $n_1$  and  $n_2$  respectively, from two distributions. By Theorem 4,

$$\begin{aligned} \sqrt{n_1} \left( \hat{G}L_{1,n_1} - GL_1 \right) &\rightarrow_d N(0, \Sigma_1), \\ \sqrt{n_2} \left( \hat{G}L_{2,n_2} - GL_2 \right) &\rightarrow_d N(0, \Sigma_2). \end{aligned}$$

Let  $n := n_1 + n_2$ . Assume that  $\lim_{n \rightarrow \infty} n_1/n = t$ . Then

$$\begin{aligned} \sqrt{n} \left( \hat{G}L_{1,n_1} - GL_1 \right) &\rightarrow_d N(0, t\Sigma_1), \\ \sqrt{n} \left( \hat{G}L_{2,n_2} - GL_2 \right) &\rightarrow_d N(0, (1-t)\Sigma_2). \end{aligned}$$

Let  $\hat{\theta}_n := \hat{G}L_{1,n_1} - \hat{G}L_{2,n_2}$ . Since  $\hat{G}L_{1,n_1}$  and  $\hat{G}L_{2,n_2}$  are independent,

$$\sqrt{n} \left( \hat{\theta}_n - \theta_0 \right) \rightarrow_d N(0, t\Sigma_1 + (1-t)\Sigma_2).$$

Thus, given  $\Sigma_1$ ,  $\Sigma_2$ , and  $t$ , we know the asymptotic distribution of  $\hat{\theta}_n$  under the least favorable case in  $H_0$ , i.e.,  $\theta_0 = 0$ .

Let  $\theta_{\min}$  be the minimum component of  $\theta_0$ . Then we can write the multivariate inequality testing problem as

$$H_0 : \theta_{\min} \geq 0 \quad \text{vs.} \quad H_1 : \theta_{\min} < 0.$$

Thus it is natural to use  $\hat{\theta}_{n,\min}$ , the minimum component of  $\hat{\theta}_n$ , as the test statistic.

Let  $F(\cdot)$  be the cdf of the minimum component of  $X \sim N(0, t\Sigma_1 + (1-t)\Sigma_2)$ . Then under the least favorable case in  $H_0$ , i.e.,  $\theta_0 = 0$ ,

$$\sqrt{n} \hat{\theta}_{n,\min} \rightarrow_d F(\cdot).$$

It is difficult to derive  $F(\cdot)$  analytically. Given  $\Sigma_1$ ,  $\Sigma_2$ , and  $t$ , however, we can draw from  $N(0, t\Sigma_1 + (1-t)\Sigma_2)$  and simulate  $F(\cdot)$ ; thus we can obtain simulated critical values. In practice, we replace  $\Sigma_1$ ,  $\Sigma_2$ , and  $t$  with their consistent estimators, and compute the asymptotic  $p$ -value by simulation.

## 5. APPLICATION: INCOME DISTRIBUTION IN JAPAN

### 5.1 Data

According to the National Survey of Family Income and Expenditure, income inequality in Japan measured by the sample Lorenz curve increased from 1979 to 1994, while income distribution measured by the sample GL curve improved during the same period (Table 1), i.e., increase in the average real income was sufficient to compensate increase in inequality. The argument based only on the point estimates, however, is incomplete.

Consider testing, for example,

$$\begin{aligned} H_0 : GL_{1994} &\geq GL_{1979} \\ \text{vs.} \quad H_1 : GL_{1994} &\not\geq GL_{1979}, \end{aligned}$$

where  $GL_{1979}$  and  $GL_{1994}$  are vectors of GL curve ordinates in 1979 and 1994 respectively. We apply the above test to this problem.

Although micro data of the National Survey of Family Income and Expenditure are not publicly available, publicly available grouped data contain sufficient information for our purpose. For each income decile group, they report the sample mean and the sample coefficient of variation of the annual income, from which we can calculate the sample second moment. Thus we can estimate the asymptotic variance-covariance matrix of the sample GL curve ordinates.

**Table 1.** The Sample GL Coordinates for the Japanese Household Real Incomes.

Decile	1979	1984	1989	1994
1	227	226	258	256
2	569	572	642	651
3	964	986	1,103	1,134
4	1,430	1,456	1,626	1,684
5	1,940	1,987	2,232	2,330
6	2,507	2,576	2,904	3,050
7	3,135	3,247	3,680	3,884
8	3,879	4,027	4,567	4,850
9	4,739	4,958	5,618	6,004
10	6,016	6,322	7,216	7,767

Note: Thousand 1995 yen deflated by the Consumer Price Index (CPI).

Source: The authors' calculation from the National Survey of Family Income and Expenditure.

## 5.2 Testing Results

The test statistic is the minimum value among the components of  $\hat{GL}_{1994} - \hat{GL}_{1979}$ , where  $\hat{GL}_{1979}$  and  $\hat{GL}_{1994}$  are vectors of the sample GL curve ordinates in 1979 and 1994 respectively. We see from Table 1 that it is 29. Given the asymptotic distribution of  $\hat{GL}_{1994} - \hat{GL}_{1979}$ , we can draw from this distribution and simulate the asymptotic distribution of the minimum component. In practice, it suffices to evaluate the asymptotic  $p$ -value rather than tabulating the distribution for each case.

Table 2 summarizes the results. For the comparison between 1979 and 1994, the  $p$ -value is 1.00; thus we accept  $H_0$ , i.e., income distribution in Japan improved from 1979 to 1994. Under the least favorable case in  $H_0$ , the test statistic is the minimum value among the multivariate normal distribution with mean 0. Unless all the components have perfect positive correlation, it has a distribution skewed to the left, i.e., it tends to be negative even under  $H_0$ . Thus

positive value of the test statistic strongly supports  $H_0$ .

**Table 2.** Testing Results.

	Test Statistic	Asy. $p$ -value
1979-1984	-0.35	0.74
1979-1989	31.16	1.00
1979-1994	29.85	1.00
1984-1989	31.51	1.00
1984-1994	30.20	1.00
1989-1994	-1.31	0.71

Note: The asymptotic  $p$ -value is based on 100,000 random draws from the asymptotic distribution of the test statistic under the least favorable case in  $H_0$ .

## 6. REFERENCES

- Andrews, D. W. K., Empirical process methods in econometrics, In: R. F. Engle and D. L. McFadden (eds), *Handbook of Econometrics*, Vol. 4, Elsevier Science, chapter 37, 1994.
- Beach, C. M. and R. Davidson, Distribution-free statistical inference with Lorenz curves and income shares, *Review of Economic Studies*, 50, 723-735, 1983.
- Berge, C., *Topological Spaces*, Dover Publications (republication in 1997), 1963.
- Dasgupta, P., A. Sen, and D. Starrett, Notes on the measurement of inequality, *Journal of Economic Theory*, 6, 180-187, 1973.
- Foster, J. E. and A. F. Shorrocks, Poverty orderings, *Econometrica*, 56, 173-177, 1988.
- Gail, M. H. and J. L. Gastwirth, A scale-free goodness-of-fit test for the exponential distribution based on the Lorenz curve, *Journal of the American Statistical Association*, 73, 787-793, 1978.
- Goldberger, A. S., One-sided and inequality tests for a pair of means, In: R. Bewley and T. van Hoa (eds), *Contributions to Consumer Demand and Econometrics: Essays in Honour of Henri Theil*, St. Martin's Press, chapter 8, 1992.
- Sen, A., *On Economic Inequality*, expanded edition, Oxford University Press, 1997.
- Shorrocks, A. F., Ranking income distributions, *Economica*, 50, 3-17, 1983.
- Yitzhaki, S. and I. Olkin, Concentration indices and concentration curves, In: K. C. Mosler and M. Scarsini (eds), *Stochastic Orders and Decision under Risk*, Institute of Mathematical Science, pp. 380-392, 1991.