THE KOREAN ECONOMIC REVIEW
Volume 22, Number 2, Winter 2006

STATISTICAL TEST OF THE REGIONAL INCOME INEQUALITY IN KOREA

EUNGWON NHO*

The paper estimates Korean gini coefficients of non-agricultural household income distribution for 16 provincial-level administrative regions and the nation by using public use micro data of National Survey of Household Income and Expenditures in 1996 and 2000, and tests hypotheses about changes of gini coefficients between the two time periods and across regions in 2000. The standard error of the gini index is estimated by jackknife and bootstrap methods rather than the traditional delta method because of the complexity of the latter.

National gini coefficient of household current income increased by 23% from 0.3257 in 1996 to 0.3934 in 2000, while changes in regional gini indices ranged from -1.7% for Jeju to +35.6% for Geonggi-do. But, the changes in gini index are not statistically significant at 10% significance level in three regions: Jeju(-1.7% from 0.3730 to 0.3667), Busan(0.74% from 0.3799 to 0.3827) and Jeonnam(4.99% from 0.3949 to 0.4146).

Regional gini indices of household total income in 2000 ranged from 0.3442 in Ulsan to 0.4528 in Gyeongbuk, but those of similar magnitude are not statistically significant. For example, Busan's income inequality(0.3852) does not differ significantly from 10 regions including Incheon(0.3614) to Daejeon(0.4107).

JEL Classification: C12, C14, D31
Keywords: gini coefficient, standard error of gini, bootstrap method, jackknife method

---

Received for publication: March 7, 2006.   Revision accepted: Nov. 9, 2006.

* Department of Economics, Chungnam National University(CNU), 220 Gung-dong, Yuseong-gu, Daejeon, 305-764, Korea. Phone: 042-821-5583, E-mail: nhoew@cnu.ac.kr. I am grateful to anonymous referees and seminar participants at the Department of Economics, Oregon State University, where I stayed as a visiting scholar in 2005, at the Joint Korean Economic Conference in 2006, and at Management and Economics Research Institute, CNU, for useful comments. Financial support from Development Fund for Chungnam National University is greatly acknowledged, Any remaining error, if any, is of course mine.

I. INTRODUCTION

Korean people are very sensitive to the inequality of income distribution. Even a slight increase in gini coefficient, a weekly magazine Hankyoreh 21(2005) reported, makes some government officials feel responsibility for it and defend themselves in one way or another. They would better direct all their energy to an improvement of real inequality if the nature of random statistics were well understood.

There's another episode. According to Dong-A Ilbo(2004), a public official of the Ministry of Finance and Economy said that the gini coefficient of the U.S. rose to 0.464 in 2003 from 0.462 in 2000 and consequently the income inequality was recently aggravated according to Census Bureau(2004a). But, the US Census Bureau(2004a,b) made it clear that "Income inequality showed *no change* between 2002 and 2003 when measured by the Gini index."(italics added). The gini index was 0.462 in 2000 as well as in 2002. The Census Bureau(2004b. p.8) adopts a 10% level as the criterion of statistical significance.

Being a statistic, any estimate regarding measures of income inequality may change due to sampling and non-sampling errors. However, even the recent researches on Korean income distribution or its changes such as Jung *et. al*.(2001), Park(2002) and Kang and Yun(2003) did not consider the variability of measures of income inequality[1]. Nor has Korean National Statistics Office(KNSO) published their standard errors yet. Thus, the paper will fill the gap by offering statistical inferences on gini coefficients in Korea.

For that matter, policy makers as well as the general public have long requested the information on their local administrative regions, especially since the activation of local governments in 1990. However, neither researchers nor KNSO has yet provided such regional inequality estimates.[2] Using the public use micro data files in 1996 and 2000

[1] A decade ago such a tendency prevailed in the U.S. As recently as in 1992, Karoly(1992) wrote that "Despite the existence of methodologies for estimating the variances of many inequality measures(Wold, 1935; Glasser, 1962; Gastwirth, 1974; Sandstrom *et. al*. 1988), many researchers do not report standard errors or discuss sampling variability.(p.108)." However, a proper and practical method has not been established yet for the gini in complex survey problems after a decade of Karoly(1992) as the following section IV shows.

[2] An exception is Nho(1999) who estimated gini coefficient, Atkinson index, mean log variation,

(explained in section II below), I will estimate gini indices and their standard errors for the 16 provincial-level local governments to test inter-regional and inter-temporal differences.

Inequality measures may be divided into two families: generalized entropy inequality measures(Theil, Atkinson index, Coefficient of Variation) and gini index(basic, extended, generalized gini). The paper focuses on basic gini index, proposed by Gini(1914), among them because of its worldwide popularity, though entropy measures are often preferred by economists in terms of the social welfare analysis and of their superior property such as decomposability.[3]

However, the characteristic of the gini index introduces complexity into estimating its standard error, that it can not be represented by functions of moments only as entropy measures can. Even the asymptotic variance estimator of gini is very complicated to be computed in the traditional delta method(Nygard and Sandstrom, 1989; Yitzahki, 1991). Recently, alternative estimation methods based on resampling have attracted more attention than the traditional one, since they do require not a derivation of complex expression for variance estimator but a computer-intensive calculation if a proper resampling method is devised. Nygard and Sandstrom(1989) and Yitzahki(1991) proposed to use the jackknife method. Mills and Zandvakili(1997) applied the standard bootstrap, and Xu(2000) applied a more efficient, iterated bootstrap, to the independently and identically distributed(iid) data, respectively. Recently, Biewen(2002) and Athanasopoulos and Vahid(2003) applied the bootstrap to the realistic data, that is, weighted survey data. The paper will apply these resampling methods with some modifications to the cases which our data structure matches well.

The paper deals with the distribution of *household income*, although welfare economists have traditionally concerned more with the distribution of individual(or personal) income. Since income data is usually collected on a household basis, the contentious concept of the equivalence scale must be posited to find an individual income[4].

---

coefficient of variation for 15 provincial-level administrative regions, but not their standard errors.

[3] See, Cowell(2000), and S.R. Chakravarty and P. Muliere(2003) for various approaches to inequality measures and their welfare implications.

[4] See, Jung *et. al*. (2002) and Park(2001, p.61) for various equivalence scales employed by

Moreover, income inequality of household has attracted more interest recently. On that matter, we compute the gini for all non-agricultural households, including single-member and multi-member households, in order to preserve the full sample size available and to allow for international comparisons.[5]

The paper is organized as follows. Section II explains the structure of our data. Section III and IV review and discuss the relevant statistical method of computing gini index and its standard error, respectively, for the complex survey data. Section V presents the results of estimations and tests of hypotheses about equalities of national gini and local gini indices overtime and across regions. Section VI summarizes our results and notes the limitations.

II. DATA

The data used was the household annual income[6] from National Survey of Household Income and Expenditures(NSHIE) in 1996 and 2000, which was conducted over all qualified[7] households by KNSO. The NSHIE is the only official data in Korea, with the sample size about 24,000, large enough to calculate regional gini coefficients, though NSHIE in 2005 was cancelled.[8]

---

researchers.

[5] It has been a Korean tradition to exclude single-member households in computing inequality measures because they were excluded from the sample of monthly Household Income and Expenditure Survey(HIES) until 2002

[6] The annual income in 1996 is summed from December 1, 1995 to November 30, 1996, while the period of income in 2000 is January 1 - December 31.

[7] The households are excluded from these surveys that are engaged in agriculture, forestry and fishery. Households with more than two employees living together, foreign households, and hospitalized patients are also excluded.

[8] There seems to be no definite criterion of the size of a region for which gini estimate is statistically reliable. However, the U.S. Census Bureau(2004c) recommends not to estimate the measures of income inequality for sub-population less than 75,000. Applying this criterion, 15 regional gini indices may be estimated because the full sample weight(number of households in population represented by a sample) of the smallest local government, Jeju Province, in 1996 is more than 93,000 households, though its sample size is 590.

[Table 1] Regional average household income(unit: 10 thousand won)

year	1996	2000				growth rate
Region	A. current income	B. current income	C. total income	D=C-B	E=D/C (%)	of current income(%)
National average	2,391.7	2,622.4	2,747.3	124.9	4.5	9.6
Seoul(SU)	2,721.3	3,092.6	3,212.4	119.7	3.7	13.6
Busan(BS)	2,486.7	2,418.8	2,501.9	83.1	3.3	-2.7
Daegu(DG)	2,292.7	2,451.5	2,630.0	178.5	6.8	6.9
Incheon(IC)	2,280.2	2,437.3	2,518.5	81.1	3.2	6.9
Gwangju(GJ)	2,321.8	2,439.9	2,581.6	141.7	5.5	5.1
Daejeon(DJ)	2,291.1	2,393.9	2,586.0	192.2	7.4	4.5
Ulsan(US)		3,033.1	3,158.9	125.8	4.0	
Gyeonggi(GG)	2,241.9	2,791.5	2,882.6	91.1	3.2	24.5
Gangwon(GW)	2,079.5	2,171.7	2,403.6	231.9	9.6	4.4
Chungbuk(CB)	2,093.6	2,255.7	2,388.1	132.4	5.5	7.7
Chungnam(CN)	2,280.1	2,203.4	2,319.5	116.1	5.0	-3.4
Jeonbuk(JB)	2,198.5	2,123.9	2,229.4	105.6	4.7	-3.4
Jeonnam(JN)	1,982.0	2,215.5	2,419.7	204.1	8.4	11.8
Gyeongbuk(GB)	1,880.4	2,272.2	2,485.6	213.4	8.6	20.8
Gyeongnam(GN)*	2,306.7	2,580.8	2,684.4	103.7	3.9	11.9
Jeju(JJ)	2,163.8	2,248.7	2,402.5	153.7	6.4	3.9
Gyeongnam(GN)		2,408.9	2,504.1	95.2	3.8	

* Ulsan is added to GN in 2000.

The household annual income is defined as the total income of all household members, which is the sum of current income and non-current income. The current income consists of earned income(wage and salary), self-employed income, realized property income and transfer income. It corresponds to the gross income, which equals market income plus public transfer, in OECD definition.(Park *et. al*. 2002, pp.25-27). Non-current income includes a retirement allowance, lump sum pension grants, benefits from congratulations and condolences, non-current subsidy, indemnity insurance money, and so on.

The KNSO reported only the annual current income in 1996, and both annual current and non-current income in 2000 HSHIE.[9] They are shown in Table 1. The household annual current income increased throughout Korea by 9.6% from 1996 to 2000, but decreased in Busan, Chungnam and Jeonbuk. The non-current income(column D in Table 1) varies from 811 to 2,319 thousand won. The total money income will be used below for testing inter-regional differences in gini coefficients in 2000. But, the current income will be used for testing an intertemporal change in gini because it is available in 1996 and 2000.

The important characteristics of the survey are summarized in the Appendix. Sample size is 24,290 households in 1996 and 23,720 in 2000. These two samples are independent because they are drawn from the 10% sampling frames of the quinquennial Population and Housing Census in 1990 and 1995, respectively. The sampling design is a stratified multi-stage sampling. The nation is partitioned to 23 strata which are allocated among 15 upper level local governments. That is, one stratum for each of 7 metropolitan cities, and two(urban and rural) strata for each of 8 Provinces. Primary Sampling Units(PSU) are selected systematically and independently within each stratum. The second stage of sampling is a random cluster sampling of 3 adjacent segments from a selected PSU. The sample is self-weighted in each stratum such that all households in a stratum have an identical design weight. Thus, the weight may be ignored in computing standard errors of the gini for each of 7 metropolitan cities, as is the case of the simple random sample(SRS) data, but not for Provinces and the nation.

The public use micro data contain stratum code, regional codes for 15 local administrative governments and normalized household design weights, but not PSU codes. This limitation makes it impossible to apply resampling methods based on PSU's or clusters. Table A in the appendix summarizes the sampling design of NSHIE.

[9] I thank an anonymous referee for pointing out the difference in the definition of the household annual income between 1996 and 2000 surveys. Non-current income was included in monthly income, but not in annual income, in 1996.

III. ESTIMATION METHODS OF GINI COEFFICIENT

1. Population gini coefficient g is defined by

$$g = \frac{d}{2\mu},$$

where d is the gini mean difference, and μ is the population mean of household income y. If population household income is assumed to be distributed by a continuous distribution function $F(y)$, then d and μ are given by

$$d = \int\int |y_1 - y_2| dF(y_1)dF(y_2), \quad \mu = \int y\, dF(y). \tag{1}$$

If a finite population $U = \{1, 2, \cdots, N\}$ is assumed, a discrete version is as follows:

$$d = \frac{1}{N(N-1)} \sum_{i=1}^{N} \sum_{j=1}^{N} |y_i - y_j|, \quad \mu = \frac{1}{N} \sum_{i=1}^{N} y_i$$

where y_i, $i = 1, 2, \cdots, N$, is the i th household's income.

Now, denote a sample by $s = \{1, 2, \cdots, n\}$. The i th unit has a pair of income y_i and sampling design weight w_i. A basic sampling design weight is $w_i^f = 1/\pi_i$, reciprocal of the selection probability π_i of the i th household, $i = 1, 2, \cdots, n$. It is called the full sample weight that means the number of population units the i th sample household represents, and hence sums to the number of total population units $\sum w_i^f = N$. (If N is unknown, it may be estimated by $\sum w_i^f$). In this paper sampling weights are normalized as $w_i = w_i^f / N$ such that $\sum_{i=1}^{n} w_i = 1$. The empirical distribution function of sample incomes is defined by

$$\hat{F}(y) = \sum_{j=1}^{n} I(y_j \le y) w_j \ ,$$

where $I(A)$ is an indicator function, taking value 1 if the condition A is true and zero otherwise. Then, the weighted sample gini coefficient is defined as

$$\hat{g} = \frac{\hat{d}}{2\hat{\mu}} \ , \tag{2}$$

where $\hat{d}$ and $\hat{\mu}$ are the same as in (1) with $F(y)$ replaced by $\hat{F}(y)$,

$$\hat{d} = \int\int |y_1 - y_2| d\hat{F}(y_1) d\hat{F}(y_2) \ , \quad \hat{\mu} = \int y \, d\hat{F}(y) = \sum_{i=1}^{n} w_i y_i \equiv \bar{y} \ .$$

It is well-known that the gini index can be interpreted as the ratio of the area above the Lorenz curve of the income distribution and below the 45° line to the area of the triangle below the 45° line.

Nygard and Sandstrom(1985, 1989) and Athanasopoulos and Vahid (2003) derived a computationally more convenient expression of the weighted gini:

$$\begin{aligned} \hat{g} &= \frac{2\sum_{j=1}^{n}\left[w_j y_{(j)} \sum_{i=1}^{j} w_i \right]}{\bar{y}} - \frac{\sum_{j=1}^{n} w_j^2 y_{(j)}}{\bar{y}} - 1 \\ &= \frac{\sum_{j=1}^{n} w_j y_{(j)} \left[2\sum_{i=1}^{j} w_i - w_j \right]}{\bar{y}} - 1 \ , \end{aligned} \tag{3}$$

where $y_{(j)}$ is the j th order statistic of y_i's, i.e., $y_{(1)} \le y_{(2)} \le \cdots \le y_{(n)}$. So did Lerman and Yitzhaki(1989) in the covariance-based equivalent formula[10]:

[10] They used the empirical distribution function defined by $\hat{F}(y_j) = \sum_{i=1}^{j-1} w_i + 0.5 w_j$. Since values

$$\hat{g} = 2\sum_{j=1}^{n} w_j (y_j - \bar{y})\left[\hat{F}(y_j) - \bar{F}\right] / \bar{y} \equiv 2\operatorname{cov}\left[y, \hat{F}(y)\right] / \bar{y}$$

where $\bar{F} = \sum_{j=1}^{n} w_j \hat{F}(y_j)$. If weights are equal, i.e., $w_i = 1/n$ for all i as in the SRS or in the data of metropolitan cities, then equation (3) is reduced to a simpler unweighted expression:

$$\hat{g} = \frac{2}{n} \frac{\sum_{j=1}^{n} j y_{(j)}}{\sum_{j=1}^{n} y_j} - \left(1 + \frac{1}{n}\right)$$

IV. ESTIMATION METHODS OF THE STANDARD ERROR OF GINI COEFFICIENT

4.1 Delta method

It is almost impossible to derive an exact distribution of gini coefficient because it is an extremely complex statistic, in particular because the gini mean difference is not a function of moments alone. So, its asymptotic property has been studied. Taking the first-order Taylor approximation of sample gini (2) round g and μ gives

$$\hat{g} - g \approx \frac{0.5\hat{d} - g\,\bar{y}}{\mu}.$$

Hoeffding(1948) derived the variance of gini coefficient for the case of iid random variables by developing a U-statistic theory[11] and proved its

---

of normalized weights are very small (< 0.0001), I used the $\hat{F}(y)$ defined above.

[11] For a sample x_i, $i = 1, 2, \cdots, n$ of size *n* and a symmetric function $h(x_1, \cdots, x_m)$ of $m(\leq n)$ variables, a U-statistic with kernel *h* and with degree *m* is defined as any function of form

$$U = U(x_1, \cdots, x_m) = \binom{n}{m}^{-1} \sum_{C_{m,n}} h(x_{i_1}, \cdots, x_{i_m})$$

asymptotic normality: [theorem 7.5(p.309)] if y_i's are identically distributed with $E[y^2] < \infty$ and $\mu = E[y] > 0$, then

$$\sqrt{n}(\hat{g} - g) \sim_a N(0, \sigma^2),$$

where

$$\sigma^2 = \frac{d^2}{4\mu^4}\zeta_1(\mu) - \frac{d}{\mu^3}\zeta_1(\mu, d) + \frac{1}{\mu^2}\zeta_1(d),$$
$$\zeta_1(\mu) = \int y^2 dF(y) - \mu^2 \equiv \sigma^2(y),$$
$$\zeta_1(\mu, d) = \int\int y_1 |y_1 - y_2| dF(y_1) dF(y_2) - \mu d,$$
$$\zeta_1(d) = \int \left[\int |y_1 - y_2| dF(y_2)\right]^2 dF(y_1) - d^2 .$$

Sendler(1979) and Nygard and Sandstrom(1981, p.384) derived a different form of gini's variance by using an L-statistic(i.e., a linear combination of the order statistics) theory. But, explicit practical expression for computing it is extremely complicated, as is shown in the appendix of Sandstrom, Wretman and Walden(1988).

For survey sample data, Nygard and Sandstrom(1985, 1989) derived three kinds of variances of gini distribution according to different approaches to sample surveys. However, the variance of gini under any approach is very complicated. Therefore, although estimation methods of the variance of gini coefficient had been developed since 1940s, "these methods were not used probably because of the complexity of the computation and computing time required."(Yitzhaki, 1991, p.235).

4.2 Jackknife method

Sandstrom, Wretman, and Walden(1988), Nygard and Sandstrom

---

where the summation is over the set $C_{m,n}$ of all $\binom{n}{m}$ combinations of m integers, $i_1 < i_2 < \cdots < i_m$ chosen from $\{1, 2, \cdots, n\}$. Sample gini mean difference is a U-statistic with $h(x_{i_1}, x_{i_2}) = |x_{i_1} - x_{i_2}|$, and so is the sample mean with $h(x_{i_1}) = x_{i_1}$.

(1989) and Yitzhaki (1991) proposed a jackknife variance estimator of the gini coefficient in the form:

$$\hat{\sigma}^2_{JK} = \frac{n-1}{n}\sum_{j=1}^{n}\left[\hat{g}^{(j)} - \hat{g}^{(\bullet)}\right]^2 \tag{4}$$

where $\hat{g}^{(j)}$ is the gini coefficient for the sample with y_j deleted, and $\hat{g}^{(\bullet)} = \sum_{j=1}^{n}\hat{g}^{(j)}/n$. These authors compared the performance of auxiliary model-based variance estimator and jackknife variance estimator by Monte Carlo studies to conclude that their sampling behaviors are quite similar to each other. But, these authors did not check the consistency of their jackknife estimators.[12] Neither did they consider complex survey samples.

The consistency of jackknife variance estimators of smooth L-statistics and their functions such as gini in complex survey problems was proved by Shao(1994). His jackknife variance estimator is defined as correcting the dependency of observations within stratum by deleting one sample cluster instead of one sample observation when dealing with complex survey problems:

$$\hat{\sigma}^2_{JKC} = \sum_{h=1}^{L}(1-f_h)\frac{n_h-1}{n_h}\sum_{i=1}^{n_h}(\hat{g}^{(hi)} - \hat{g})^2,$$

where $f_h = n_h / N_h =$ sampling rate of stratum h, N_h and n_h are the number of clusters in stratum h of population and of the sample, respectively, $\hat{g}^{(hi)}$ is the gini coefficient for the sample with the ith cluster in stratum h that is deleted. Such a estimator may not be feasible in practice because the information on cluster is often not open to the public, as is the case with my data.

[12] Yitzhaki(1991) refers the applicability of jackknife method to gini to the fact that gini is a U-statistic and to Efron(1982, p.26). But, it is not clear how this method can be justified to such a complex statistic as gini coefficient from the complex survey samples.

4.3 Bootstrap method

The standard bootstrap method, pioneered by Efron(1979), is to draw a large number of bootstrap samples $\{y_{b1}^*, y_{b2}^*, \cdots, y_{bn}^*\}, b = 1,2,\cdots, B_1$ of size n from the original, iid sample $\{y_1, y_2, \cdots, y_n\}$ with replacement, and to compute the statistic of interest, such as $\hat{g}_b^*$, $b = 1,2,\cdots, B_1$. Then, a standard bootstrap variance estimator of parameter g is given by

$$\hat{\sigma}_{BS}^2 = \frac{1}{B_1 - 1}\sum_{b=1}^{B_1}(\hat{g}_b^* - \hat{g})^2 .$$

Also, the distribution of the estimator is estimated by

$$\hat{H}_{BS}(x) = \frac{1}{B_1}\sum_{b=1}^{B_1} I\{\hat{g}_b^* \le x\}$$

Its percentiles can be used to construct the $(1-\alpha)$-level percentile confidence interval $(\hat{g}_L, \hat{g}_H)$ where

$$\alpha/2 = \hat{H}_{BS}(\hat{g}_L),\ \ 1-\alpha/2 = \hat{H}_{BS}(\hat{g}_H) .$$

This is known as the percentile method.

Mills and Zandvakili(1997) seem to be the first to apply the standard bootstrap method to estimate the standard error of gini coefficient of income distribution, ignoring sampling weights of the survey data.[13] They found that asymptotic standard errors were very similar for a large sample size whether calculated by delta method or by standard bootstrap method. Recently, Moran(2005) also applied it to the income survey of the Luxemburg Income Study(LIS) database.

Bickel and Freedman(1981) showed that the bootstrap U-statistic estimator is consistent if expectation of the absolute value of the kernel and that of squared kernel are finite. Since gini coefficient is a function of U-statistics with finite expectations of the kernels, its bootstrap estimator

[13] An ecologist Weiner(1985) applied the standard bootstrap method to the gini coefficient of the distribution of plant size as early as in 1985.

is consistent.

The appealing merit of bootstrap estimators does not lie in their consistency but in their accuracy. Hall(1992) showed that the coverage error of the bootstrap-t confidence interval based on double(or two-level nested) bootstrapping is of order $O_p(n^{-1})$ while that of the percentile confidence interval based on the standard bootstrap or standard normal approximation is $O_p(n^{-1/2})$.

The bootstrap-t confidence interval for a parameter θ is computed by percentiles of distribution

$$\hat{G}_{BS}(x) = \frac{1}{B_1}\sum_{b=1}^{B_1} I\{T_b^* \le x\} \tag{5}$$

of a studentized statistic of the bootstrapped estimators $\hat{\theta}_b^*$:

$$T_b^* = \frac{\hat{\theta}_b^* - \hat{\theta}}{\hat{\sigma}_b^*}, \tag{6}$$

where $\hat{\sigma}_b^*$ is the estimate of the standard error of $\hat{\theta}_b^*$ (Moran, 2005). If $\hat{\sigma}_b^*$ is not computed from the bth resample { $y_{b1}^*, y_{b2}^*, \cdots, y_{bn}^*$ }, B_2 second-round resamples { $y_{b1}^{**}, y_{b2}^{**}, \cdots, y_{bn}^{**}$ } must be drawn from the former in the same manner as the first-round resample, and can be used to compute the standard error of $\hat{\theta}_b^*$. This double bootstrap incurs a heavy load of computation, i.e., $B_1 \times B_2$ number of computing $\hat{\theta}_b^*$'s. Then, a bootstrap-t two-sided confidence interval with $1-2\alpha$ level is given by $(\hat{\theta}_L, \hat{\theta}_H)$ where

$$\hat{\theta}_L = \hat{\theta} - \hat{\sigma}_{BS}\hat{G}_{BS}^{-1}(1-\alpha), \quad \hat{\theta}_H = \hat{\theta} - \hat{\sigma}_{BS}\hat{G}_{BS}^{-1}(\alpha). \tag{7}$$

Xu(2000) used a double bootstrap method to improve the accuracy of the statistical inference for generalized gini indices.

A straightforward extension of the bootstrap to survey problems is the naive bootstrap which applies the standard bootstrap in each stratum. It draws a SRS bootstrap sample of n_h PSU's(or clusters) with replacement from the original n_h PSU's in each stratum h, and each ultimate observation unit carries its own sampling weight(Shao and Tu, 1995, p.246).

The naive bootstrap may yield an inconsistent estimator. One source of the inconsistency is the finite population. That is, if the original sample is taken without replacement while bootstrap sampling is with replacement, the empirical distribution of a bootstrap estimator does not provide an asymptotically valid approximation to the distribution of the original estimator unless sampling ratios approaches zero. A remedy is the with-replacement bootstrap. It is to take a SRS sample of size m_h, instead of n_h, clusters with replacement from the original sample clusters in each stratum h: $m_h = (n_h - 1)/(1 - f_h)$, $f_h =$ sampling ratio of stratum h.(Shao and Tu, 1995, pp.247-8).

However, the naive bootstrap and its remedy are often hardly applicable in practice due to the lack of PSU information in public use micro data set. Recently, Biewen(2002) and Athanasopoulos and Vahid(2003) applied the bootstrap to the inference on inequality measures for weighted survey data, but drew a bootstrap sample of the ultimate sampling units, not of the PSU's, coupled with their sampling weights directly from the original sample. Moreover, they did not consider stratification.

Biewen(2002) showed by Monte Carlo experiments that the simple bootstrap provides as good a confidence interval as not only the conventional normal approximation but also a more advanced studentized bootstrap-t for realistic population distribution and sample sizes.[14] [15]

In this paper, I will apply the bootstrap to each stratum independently as in the naive bootstrap because my data is stratified. But, bootstrapping

[14] Biewen(2002) focuses on the inequality measures such as Theil coefficient and Atkinson index which are a function of moments only such that he could prove the consistency of their bootstrap estimators. Gini index is not a member of such measures and was excluded from his analysis.

[15] Athanasopoulos and Vahid(2003) also tried another bootstrap method; draw a bootstrap sample y_j^*, $j = 1, 2, \cdots, n$ from the empirical distribution

$$\hat{F}(x) = n^{-1} \sum_{i=1}^{n} w_i I\{y_i \le x\}, \quad \sum_{i=1}^{n} w_i = 1,$$

and then compute the unweighted gini. It may be regarded as a version of Gross(1980) population bootstrap. These authors reported that, for gini and Theil indices, their method produced smaller standard errors than Biewen(2002), but not small enough to affect conclusions based on the latter.

is applied to the ultimate sampling units(individual households), not to PSU's, as in Biewen(2002) and Athanasopoulos and Vahid(2003) because PSU codes are not available.

V. ESTIMATION AND TEST OF REGIONAL GINI'S IN KOREA

5.1 National gini

Korean national gini coefficients are shown in Table 2. The gini estimated by Lerman and Yitzhaki(1989)'s covariance approach is 0.3257 in 1996 and increased by 20.8% to 0.3934 in 2000 as shown in row a. Nygard and Sandstrom(1985, 1989)'s formula (3) based on L-statistic produces almost the same estimates.[16]

The 20.8% change in sample gini values between 1996 and 2000 is so large that a blind guess suggests a change in population gini. But, is it really statistically significant? To test the null hypothesis $H_0 : dg = g_{2000} - g_{1996} = 0$ against $dg \neq 0$, test statistic is computed as follows:

$$t = \frac{\hat{dg}}{se(\hat{dg})} = \frac{\hat{g}_{2000} - \hat{g}_{1996}}{\sqrt{\hat{\sigma}^2(\hat{g}_{1996}) + \hat{\sigma}^2(\hat{g}_{2000})}}, \tag{8}$$

where $se(\hat{dg})$ is the standard error of $\hat{dg} = \hat{g}_{2000} - \hat{g}_{1996}$. Row c in Table 2 shows the jackknife standard errors and t-values.Row d shows the bootstrap standard errors and t-values, where 1,000(= B_1) resamples were drawn independently for each stratum in each year.[17] These t-values prove the statistical significance of the change in Korean national gini.

Since the PSU's can not be identified in our data, jackknife variance estimator (4) was applied as a delete-household, not as a delete-cluster, type. And the naive bootstrap method was also applied to the ultimate units(households). Thus, a problem with the naive bootstrap with stratification as noted above does not arises in this modified naive

[16] The gini coefficients published by KNSO are 0.2950 in 1996 and 0.3620 in 2000, increasing 22.7%. But, they are calculated only for non-single households.

[17] Bootstrap resamples were drawn by Proc SurveySelect in SAS program. And all computing in this paper was made by SAS 9.1.

bootstrap. It is interesting that the jackknife standard error is lower than that of the bootstrap in 1996, but higher in 2001, though the differences are within ±2.5%.

[Table 2] National gini coefficients and test of their change in Korea

	method	1996	2000	growth or t-value
Gini coefficient	a. Covariance app	0.32566	0.39342	growth 20.8%
	b. L-statistic app	0.32565	0.39342	growth 20.8%
standard error of gini	c. Jackknife	0.00245	0.00537	t-value = 11.85
	d. bootstrap	0.00239	0.00554	t-value = 11.67

5.2 Regional gini's

Regional gini coefficients and their bootstrap standard errors are reported at Table 3. Since each metropolitan city forms a single stratum, all households in the city have an identical design weight, which may be ignored in computing the bootstrap standard error. For 8 Provinces, the bootstrap samples were drawn from each of their two strata separately on the level of household.

First, consider the inter-temporal changes in regional gini's. The test statistic is similar to (8) above. The 7th column of Table 3 shows the percentage rate of change in regional gini's. Geonggi leads the rising trend with 35.6% growth rate while Jeju is the only exception to this trend with 1.7% decrease. The regions are apparent from Figure 1 that have relatively lower rates of change, as indicated by arrows: Jeju, Busan and Jeonnam. As the last column of Table 3 shows, the changes in regional gini's in these three areas are not statistically significant as they do not meet the 10% level of significance. Thus, one may conclude that income inequality of households measured by gini index did not change in Jeju, Busan and Jeonnam Provinces during 1996-2000 period, while the national gini increased dramatically.

Note that Jeonbuk Province with t-value 1.95 is the only region whose change in gini is statistically significant by meeting the 10% level of significance with critical value 1.645, but barely not at the 5% level with critical value 1.96. It may be conjectured that this significance results from an inaccurate normal approximation of asymptotic distribution or the percentile confidence interval based on the standard bootstrap. Thus, it is worthwhile to try the bootstrap-t test that is based on double bootstrap method, and hence more efficient than the standard test.

[Table 3] Regional gini coefficients and the tests of change

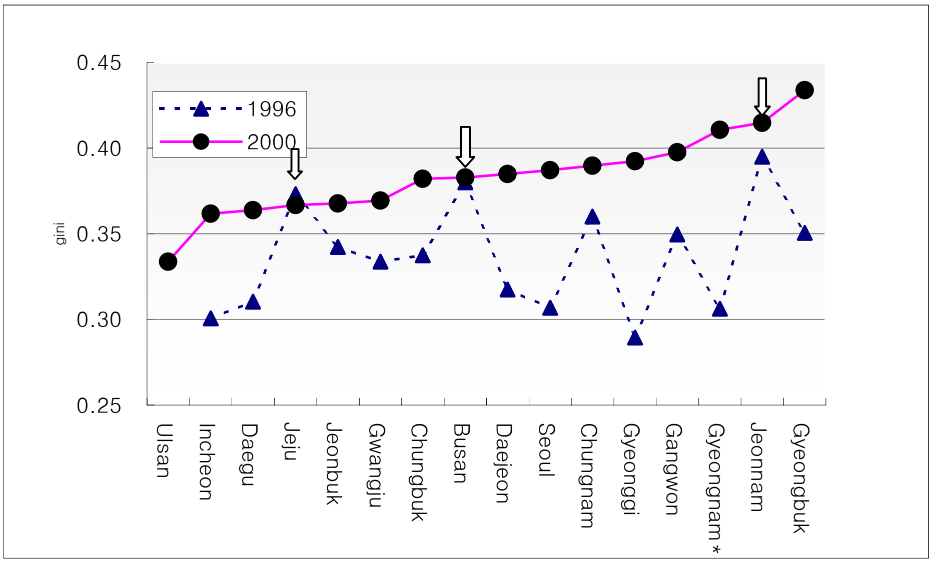
	1996		2000		test		
region	gini	se	gini	se	dif-ference	% change	t-value
Seoul(SU)	0.3069	0.0044	0.3870	0.0140	0.0801	26.12	5.47*
Busan(BS)	0.3799	0.0137	0.3827	0.0154	0.0028	0.74	0.14
Daegu(DG)	0.3103	0.0060	0.3637	0.0073	0.0534	17.21	5.66*
Incheon(IC)	0.3006	0.0067	0.3618	0.0083	0.0611	20.33	5.71*
Gwangju(GJ)	0.3336	0.0057	0.3694	0.0088	0.0358	10.72	3.42*
Daejeon(DJ)	0.3174	0.0051	0.3848	0.0154	0.0674	21.24	4.15*
Ulsan(US)			0.3336	0.0157			
Gyeonggi(GG)	0.2895	0.0053	0.3924	0.0137	0.1029	35.56	7.00*
Gangwon(GW)	0.3495	0.0075	0.3975	0.0097	0.0480	13.74	3.92*
Chungbuk(CB)	0.3374	0.0088	0.3820	0.0099	0.0445	13.20	3.37*
Chungnam(CN)	0.3600	0.0109	0.3896	0.0103	0.0296	8.21	1.98*
Jeonbuk(JB)	0.3422	0.0103	0.3677	0.0080	0.0255	7.45	1.95*
Jeonnam(JN)	0.3949	0.0095	0.4146	0.0166	0.0197	4.99	1.03
Gyeongbuk(GB)	0.3505	0.0077	0.4338	0.0151	0.0833	23.76	4.91*
Gyeongnam(GN)+	0.3063	0.0056	0.4106	0.0191	0.1043	34.04	5.25*
Jeju(JJ)	0.3730	0.0125	0.3667	0.0103	-0.0064	-1.71	-0.39
Gyeongnam(GN)++			0.4352	0.0270			

+ Ulsan is added in 2000, ++ Ulsan is excluded.

* statistically significant at the 10% level of significance

Let $\hat{\theta}$ be $\hat{dg}_{jb} = \hat{g}_{jb.2000} - \hat{g}_{jb.1996}$ in (6), where subscript jb denotes Jeonbuk. I drew 200($= B_2$) resamples from each resample $b = 1,\cdots,1000$ of Jeonbuk in 1996 and 2000 independently for each stratum, and computed $\hat{\sigma}_{BS}(\hat{dg}^*_{jb.b})$ and T^*_b, $b = 1,\cdots,1000$. By (5), 5% and 95% percentiles are -1.5819017 and 1.6959828, respectively. Then from (7), the 90% bootstrap-t confidence interval is computed as [0.003342, 0.046141]. Since it does not include zero, the null hypothesis $dg_{jb} = g_{jb.2000} - g_{jb.1996} = 0$ can be rejected at 10% significance level. On the other hand, since 2.5% percentile is calculated as -1.90104 and 97.5% percentile as 2.14409, the 95% bootstrap-t confidence interval is given as [-0.00251, 0.05031]. Since it includes zero, the null hypothesis $dg_{jb} = 0$ can not be rejected. Thus, the result for Jeonbuk based on the standard bootstrap does not change by the double bootstrap method.[18]

[Figure 1] Regional gini's in 1996 and 2000



Secondly, look at inter-regional differences of regional gini's in 2000. Now, annual "total" household income including non-current income is more relevant than annual current income for the analysis of income distribution. Table 4 shows regional gini indices and their standard errors for the two kinds of incomes. Adding non-current income to current income turns out to raise income inequality measured by gini except for two regions(Incheon and Jeju)[19], and its effect on the standard error of gini is much larger, ranging from -9.7% to 23.1%.

[18] The bootstrap-t test for other regions was not undertaken due to its high cost in terms of computing time.

[19] Statistical tests for these statements were not undertaken.

[Table 4] regional gini's of current and total incomes in 2000

	Current income		total income			rate of change	
	gini	se	gini	%	se	gini	se
Ulsan	0.3336	0.0157	0.3442	100.0	0.0150	3.18	-4.50
Incheon	0.3618	0.0083	0.3614	105.0	0.0081	-0.10	-2.95
Jeju	0.3667	0.0103	0.3634	105.6	0.0100	-0.88	-2.71
Daegu	0.3637	0.0073	0.3694	107.3	0.0079	1.57	8.37
Gwangju	0.3694	0.0088	0.3757	109.2	0.0095	1.71	7.89
Jeonbuk	0.3677	0.0080	0.3760	109.2	0.0098	2.24	21.95
Busan	0.3827	0.0154	0.3852	111.9	0.0150	0.66	-3.08
Chungbuk	0.3820	0.0099	0.3872	112.5	0.0101	1.38	1.67
Chungnam	0.3896	0.0103	0.3949	114.7	0.0110	1.36	7.33
Gyeonggi	0.3924	0.0137	0.3954	114.9	0.0134	0.76	-2.57
Seoul	0.3870	0.0140	0.3970	115.4	0.0133	2.60	-4.74
Daejeon	0.3848	0.0154	0.4107	119.3	0.0158	6.73	2.66
Jeonnam	0.4146	0.0166	0.4211	122.3	0.0150	1.56	-9.66
Gangwon	0.3975	0.0097	0.4273	124.2	0.0119	7.50	23.14
Gyeongnam	0.4352	0.0270	0.4377	127.2	0.0263	0.57	-2.47
Gyeongbuk	0.4338	0.0151	0.4528	131.5	0.0151	4.37	-0.36

In table 4, regions are arranged in the ascending order of gini indices of total income in 2000. Gyeongbuk has the highest gini(0.4528) which is 31.5% higher than the lowest gini 0.3442 of Ulsan. But, does Gyeongbuk's gini really differ from the second highest gini 0.4377 of Gyeongnam or the sixth highest gini 0.3970 of Seoul?

To answer these questions, t-test for the difference between any two regions is undertaken. The test statistic is similar to (8), and t-values are presented in Table 5. Here, boldfaced numbers indicate statistical significance at the 10% level of significance. For example, the last row indicates that income inequality in Gyeongbuk(0.4528) does differ from that of Daejeon(0.4107) significantly, but not from Jeonnam(0.4211). Taking another example, the eighth rows states that Busan's income inequality(0.3852) does not differ significantly from Incheon's(0.3614) to Daejeon(0.4107). Table 5 shows that in general, a region's income inequality does not significantly differ from those of the adjacent regions which lie within third or fifth rankings upward or downward.

[Table 5] t-values of the tests of regional differences in gini's in 2000(total income)

	US	IC	JJ	DG	GJ	JB	BS	CB	CN	GG	SU	DJ	JN	GW	GN	GB
US		1.01	1.07	1.49	**1.78**	**1.78**	**1.94**	**2.39**	**2.73**	**2.55**	**2.64**	**3.05**	**3.63**	**4.34**	**3.09**	**5.11**
IC	1.01		0.16	0.71	1.15	1.15	1.40	**2.00**	**2.45**	**2.18**	**2.29**	**2.77**	**3.51**	**4.58**	**2.77**	**5.34**
JJ	1.07	0.16		0.47	0.89	0.90	1.21	**1.68**	**2.11**	**1.91**	**2.02**	**2.52**	**3.20**	**4.11**	**2.64**	**4.94**
DG	1.49	0.71	0.47		0.51	0.52	0.93	1.39	**1.88**	**1.67**	**1.78**	**2.33**	**3.05**	**4.04**	**2.48**	**4.89**
GJ	**1.78**	1.15	0.89	0.51		0.02	0.54	0.83	1.32	1.20	1.30	**1.89**	**2.56**	**3.39**	**2.22**	**4.32**
JB	**1.78**	1.15	0.90	0.52	0.02		0.52	0.80	1.28	1.17	1.27	**1.87**	**2.52**	**3.33**	**2.20**	**4.27**
BS	**1.94**	1.40	1.21	0.93	0.54	0.52		0.11	0.52	0.51	0.59	1.17	**1.69**	**2.20**	**1.73**	**3.18**
CB	**2.39**	**2.00**	**1.68**	1.39	0.83	0.80	0.11		0.51	0.49	0.59	1.25	**1.88**	**2.57**	**1.79**	**3.62**
CN	**2.73**	**2.45**	**2.11**	**1.88**	1.32	1.28	0.52	0.51		0.03	0.12	0.82	1.41	**2.00**	1.50	**3.10**
GG	**2.55**	**2.18**	**1.91**	**1.67**	1.20	1.17	0.51	0.49	0.03		0.09	0.74	1.28	**1.78**	1.43	**2.85**
SU	**2.64**	**2.29**	**2.02**	**1.78**	1.30	1.27	0.59	0.59	0.12	0.09		0.66	1.20	**1.70**	1.38	**2.77**
DJ	**3.05**	**2.77**	**2.52**	**2.33**	**1.89**	**1.87**	1.17	1.25	0.82	0.74	0.66		0.48	0.84	0.88	**1.93**
JN	**3.63**	**3.51**	**3.20**	**3.05**	**2.56**	**2.52**	**1.69**	**1.88**	1.41	1.28	1.20	0.48		0.33	0.55	1.49
GW	**4.34**	**4.58**	**4.11**	**4.04**	**3.39**	**3.33**	**2.20**	**2.57**	**2.00**	**1.78**	**1.70**	0.84	0.33		0.36	1.32
GN	**3.09**	**2.77**	**2.64**	**2.48**	**2.22**	**2.20**	**1.73**	**1.79**	1.50	1.43	1.38	0.88	0.55	0.36		0.50
GB	**5.11**	**5.34**	**4.94**	**4.89**	**4.32**	**4.27**	**3.18**	**3.62**	**3.10**	**2.85**	**2.77**	**1.93**	1.49	1.32	0.50	

* Boldfaced numbers denote statistical significancy at the 10% level of significance(t-value less than 1.645).

VI. CONCLUSION

The paper estimated Korean gini coefficients of household income distribution for 16 Province-level regions and the nation by using public use micro data of National Survey of Household Income and Expenditures in 1996 and 2000, and tested hypothesis about changes of gini's between the two time periods and across regions in 2000. Household current income was used for intertemporal tests while household total income(non- current income added) for interregional tests in 2006. The standard errors of the gini's were estimated by the jackknife and the bootstrap methods rather than the traditional delta method because of the complexity of the latter.

Between 1996 and 2000, national gini increased by 20.8% while regional gini's changed from -1.7% to +35.6%. All the changes were found to be statistically significant at 10% significance level except for three regions - Busan, Jeju and Jeonnam Provinces. A more efficient double bootstrap test was applied to Jeonbook, but could not change the conclusion based on the standard bootstrap.

When regional gini's in 2000 are arranged in the ascending order, the adjacent regions within the third or fifth rankings upward or downward from any region do not show significant differences.

The resampling methods applied in this paper are similar to those of its predecessors. But, their statistical properties need be further explored. For example, I applied the jackknife to individual observation units rather than the primary sampling units(PSUs) and also the naive bootstrap to individual observation units. Though some results from Monte Carlo experiments suggest the relevance of those methods, a rigorous proof is yet to be established, because, as Shao and Tu(1995, p.17) warn, blind application of these methods may lead to incorrect results in complex survey problems.

Appendix [Table A]. Sampling design of National Survey of Household Income and Expenditures (NSHIE) in 1996 and 2000

		1996		2000	
region	strata*	sample size	design weight+	sample size	design weight
Seoul Metropolitan(SU)	0	3,111	9.7773E-05	3,517	6.72605E-05
Busan Metropolitan(BS)	0	2,539	3.75219E-05	2,017	4.22996E-05
Daegu Metropolitan(DG)	0	1,712	3.13632E-05	1,494	3.83298E-05
Incheon Metropolitan(IC)	0	2,144	2.55168E-05	2,049	2.82176E-05
Gwangju Metropolitan(GJ)	0	1,710	1.60711E-05	1,753	1.76028E-05
Daejeon Metropolitan(DJ)	0	1,669	1.48324E-05	1,419	2.22289E-05
Ulsan Metropolitan(US)	0			880	2.61653E-05
Gyeonggi-do(GG)	1	1,915	6.37529E-05	1,621	0.000100775
	2	816	5.31789E-05	802	4.16245E-05
Gangwon-do(GW)	1	782	2.40552E-05	754	2.84382E-05
	2	442	2.4776E-05	513	2.02836E-05
Chungcheongbuk-do(CB)	1	776	1.95738E-05	635	3.13083E-05
	2	354	2.05068E-05	363	2.21531E-05
Chungcheongnam-do(CN)	1	636	1.44811E-05	611	2.22523E-05
	2	386	3.69655E-05	474	3.91153E-05
Jeollabuk-do(JB)	1	751	3.04557E-05	687	4.09292E-05
	2	201	3.25065E-05	243	2.76115E-05
Jeollanam-do(JN)	1	609	2.51346E-05	769	2.44205E-05
	2	326	4.13279E-05	290	5.21757E-05
Gyeongsangbuk-do(GB)	1	698	3.65408E-05	651	4.6148E-05
	2	454	4.43722E-05	466	4.32356E-05
Gyeongsangnam-do(GN)++	1	1,223	5.17173E-05	740	5.58792E-05
	2	446	5.25294E-05	393	4.89635E-05
Jeju-do(JJ)	1	413	1.46753E-05	410	1.80708E-05
	2	177	1.22842E-05	169	1.20777E-05
total		24,290		23,720	

* Statum 1 denotes urban(city) area and stratum 2 rural(county) area.

\+ Design weights are individual household's weight, and those of 2000 are rescaled so as for their sum to be unity.

++ GN Province includes Ulsan Metropolitan in 1996, which was designated as metropolitan in 1997.

Reference

Athanasopoulos, G. and F. Vahid (2003), “Statistical inference and changes in income inequality in Australia,” *The Economic Record*, Vol. 75, 412-414.

Biewen, M. (2002), “Bootstrap inference for inequality, mobility and poverty measurement,” *Journal of Econometrics*, Vol. 108, 317-342.

Bickel, P. J. and D. A. Freedman (1981), “Some asymptotic theory for the bootstrap,” *The Annals of Statistics*, Vol. 9, 1196-1217.

Chakravarty, S. R. and P. Muliere (2003), “Welfare indicators: a review and new perspectives. 1. Measurement of inequality,” *Metron*, Vol. LXI, No. 3, 457-497.

Cowell, F. (2000), “Measurement of Inequality,” in A.B.Atkinson and F. Bourguignon (eds) *Handbook of Income Distribution*, North-holland.

The Dong-A Ilbo (2004), “MOFE: income inequality severe in America than in Korea,” Sept. 6.

Efron, B. (1979), “Bootstrap methods: another look at the jackknife,” *Annals of Statistics*, Vol. 7, 1-26.

Efron, B. (1982), *The Jackknife, the bootstrap and other resampling plans*, The society for industrial and applied mathematics, Philadelphia, Pennsylvania.

Gini, C. (1914), “On the measurement of concentration and variability of characters,” (Italian) *Atti del Reale Istituto Veneto di Scienze, Lettere ed Arti*, Vol. LXXIII- Part II, 1203-1248. English translation of this paper was published recently by Metron: Gini, C. (2005), “On the measurement of concentration and variability of characters,” *Metron*, Vol. LXIII, n.1, 3-38.

Gross, S. T. (1980), “Median Estimation In Sample Surveys,” *Proceedings of the section on survey research methods*, 81-184, American Statistical Association.

Hall, P. (1992), *The Bootstrap and Edgeworth Expansion*. Springer, New York.

Hankyoreh 21 (2005), “Rampant Statistics: 5.87 times,” May 24.

Hoeffding, W. (1948), “A class of statistics with asymptotically normal distribution,” *Annals of Mathematical Statistics*, Vol. 19, 293-325.

Jung, Jin Ho, et. al. (2001), *The state and policy issues of income distribution and poverty*(in Korean), Korea Labor Institute.

Kang, Byung-Goo and Myeong-Su Yun (2003), “Decomposing labor income inequality change,” *Journal of Korean Economic Development*, Vol. 9, No. 1, 155-175.

Karoly, L.A. (1992), "Changes in the Distribution of Individual Earnings in the United States-1967-1986," *The Review of Economics and Statistics*, Vol. 74, No. 1, 107-115.

Lerman, R. and S. Yitzhaki (1989), "Improving the accuracy of estimates of Gini coefficients," *Journal of Econometrics*, Vol. 42, 43-47.

Mills, A. M. and S. Zandvakili (1997), "Statistical inference via bootstrapping for measures of inequality," *Journal of Applied Econometrics*, Vol. 12, 133-150.

Moran, T. P. (2005), "Bootstrapping the LIS: Statistical Inference with the Gini Index, and Patterns of Inequality in the Global North," Luxemburg Income Study Working Paper, No. 378.

Nho, Eungwon (1999), "The inequalities and its decomposition of regional household income distributions in Korea," (in Korean) Chungnam National University, *Economic Review*, Vol. 15, 111-124

Nygard, F. and A. Sandstrom (1981), *Measuring Income Inequality*. Almqvist & Wiksell, Stockholm.

Nygard, F. and A. Sandstrom (1985), "The estimation of the gini and the entropy inequality parameters in finite populations," *Journal of Official Statistics*, Vol. 4, No. 1, 399-412.

Nygard, F. and A. Sandstrom (1989), "Income inequality measures based on sample surveys," *Journal of Econometrics*, Vol. 42, 81-95.

Park, Chan Yong, et. al. (2002), *Analysis of the movements, and factors of change, in income distribution and poverty*(in Korean), Korea Institute for Health and Social Affairs.

Sandstrom, A., Wretman, J. H. and B. Walden (1988), "Variance estimators of the gini coefficient−probability sampling," *Journal of Business and Economic Statistics*, Vol. 6, No. 1, 113-119.

Shao, J. (1994), "L-statistics in complex survey problems," *The Annals of Statistics*, Vol. 22, No. 2, 946-967.

Shao, J. and D. Tu (1995), *The Jackknife and the Bootstrap*. Springer, New York.

Sendler, W. (1979), "On statistical inference in concentration measurement," *Metrika*, Vol. 26, 109-122.

US Census Bureau (2004a), "Income Stable, Poverty Up, Numbers of Americans With and Without Health Insurance Rise," Census Bureau Reports (Press Release), August 26, http://www.census.gov/Press-Release/www/releases/archives/income_wealth/002484.html)

US Census Bureau (2004b), *Income, poverty, and health insurance coverage in the United States: 2003,* p.1.

US Census Bureau (2004c), "Table HINC-01. Selected Characteristics of Households, by Total Money Income in 2001," (http://ferret.bls.census.gov/macro/032002/hhinc/new01_000.htm.)

Weiner, J. (1985), "Size hierarchies in experimental populations of annual plants," *Ecology*, Vol. 66, No. 3, 743-752.

Xu, K. (2000), "Inference for generalized gini indices using the iterated-bootstrap method," *Journal of Business and Economic Statistics*, Vol. 18, No. 2, 223-222.

Yitzahki, S. (1991), "Calculating jackknife variance estimators for parameters of the gini method," *Journal of Business and Economic Statistics*, Vol. 9, No. 2, 235-239.

Yitzahki, S. (1998), "More than a dozen alternativ ways of spelling gini," *Research on Economic Inequality*, Vol. 8, 13-30.