Jour. Ind. Soc. Ag. Statistics
Vol. XLII, No. 1 (1990), pp. 118-130

TESTING THE SIGNIFICANCE RELATED TO GINI RATIO, NON-PARAMETRIC TEST STATISTICS

S. ARORA, A. A. JULKA and O. P. BAGAI
Panjab University, Chandigarh-160014
(Received : September, 1988)

SUMMARY

Using convergence in probability and Slutsky's Theorem, asymptotic distribution of Gini-ratio is obtained. The derivation covers the case of mixed (negative as well as positive) income profile.

Keywords : Inequality ; Gini-ratio ; Normalisation ; Negative Income ; Convergence in Probability.

1. Introduction

Ever since its advent in 1912, the Gini coefficient or Gini ratio continues to be in vogue and is practitioner's favourite, particularly in empirical studies on income inequality. When the distribution profile has only positive incomes, the asymptotic distribution of estimated Gini ratio is given by Ramakrishnan [8]. The related test of significance for either a specified Gini ratio or for the difference between two Gini ratios, can easily be constructed. But in case the distribution profile has both negative [1] and positive incomes — the case of negative incomes

1. Those dealing with income distribution aspects of rural economy in developing economies will appreciate that we do encounter negative incomes, farm income variously defined and profit etc. Also the phenomenon of negative incomes (i.e. losses) is so frequent in industries ; see, for instance Hagerbaumer [5], Pyatt-chen Fei [7], Singh, Ajit [11]. Studies in Economics of Farm Management, various states, Issued by :— The Directorate of Economics and Statistics, Ministry of Food and Agriculture, Govt. of India.

being visualised in the losses which a farm firm or an organisation may have, the sampling distribution of Gini ratio is still not known.

This paper is an attempt to plug this very gap by finding non-parametric test statistic and its asymptotic distribution. Further, the usual definition of Gini ratio would require certain modifications.[2] However, Chen *et al.* [2] have suggested a normalised Gini ratio, in case the distribution includes both positive and negative entries, under the condition that the mean of the distribution is positive. The modified measure retains all the basic properties of the conventional Gini ratio. We have followed this definition of normalised Gini ratio along with its two versions (see Chen *et al.* [2] which have been defined in Section 2. Once the asymptotic distribution of the statistic, for both the versions of normalised Gini ratio, is obtained the related tests of significance based on single sample or two independent samples can easily be developed; as suggested in Section 3.

2. Definitions and Preliminaries

Let $Y_1, Y_2, \ldots, Y_n$ be a random sample of size n, drawn from an unknown income distribution $F(Y)$. It is assumed that the distribution has a finite mean μ and has a differentiable density. We also assume that some of the incomes, Y_i's might be negative but the total income is positive, i.e.

$$\sum_{i=1}^{n} Y_i > 0 \tag{2.1}$$

It might be added that the case of $\sum_{i=1}^{n} Y_i < 0$ might throw further theoretical possibilities but this makes an economic absurdity. Hence in this paper we restrict ourselves to the realistic case of a viable economy producing positive income in the aggregate. Let the incomes be ordered as

$$Y_1 \leqslant Y_2 \leqslant \ldots \leqslant Y_n$$

2. In case, the data profile has both negative and positive entries, the Gini ratio may exceed one and thus may overestimate the inequality. To avoid this. some adjustment is needed in this case. Till recently the usual procedure was to ignore negative entries by taking them as zeros and then compute Gini ratio.

Then the Gini ratio is estimated as

$$\hat{G} = \frac{\hat{\Delta}}{2\,\bar{Y}} \tag{2.2}$$

where $$\hat{\Delta} = \frac{2}{n\,(n-1)} \sum_{i=1}^{n} \sum_{j<i} (Y_i - Y_j) \tag{2.3}$$

and $$\bar{Y} = \frac{1}{n} \sum_{i=1}^{n} Y_i \tag{2.4}$$

Let $$y_i = \frac{Y_i}{n\,\bar{Y}}, \quad i = 1 \ldots n \tag{2.5}$$

where y_i is the income share of the ith unit.

Let us define k as the largest integer such that

$$\sum_{i=1}^{k} y_i = 0 \tag{2.6}$$

Following Sen [9] and Chen *et al.* [2] and using the two definitions (2.5) and (2.6), the expression for $\hat{G}$ can be written as follows :

$$\hat{G} = \frac{1}{n-1} \sum_{i=1}^{n} \sum_{j<i}^{n} (y_i - y_j)$$

$$= \frac{1}{n-1}\left[\sum_{i=1}^{n} y_i\,(2i - (n+1))\right]$$

$$= \frac{1}{n-1}\left[\sum_{i=1}^{n} y_i\,(n - (1 + 2\,(n-i))\right]$$

$$= 1 + \frac{1}{n-1} - \frac{1}{n-1} \sum_{i=1}^{n} y_i\,(1 + 2\,(n-i))$$

$$= 1 + \frac{1}{n-1} + \frac{2}{n-1} \left(\sum_{i=1}^{k} i\, y_i \right)$$

$$- \frac{1}{n-1} \sum_{i=k+1}^{n} y_i \, (1 + 2\,(n-i)) \qquad (2.7)$$

A formulation as in (2.7) brings home the point that the coefficient can take a value greater than one, when negative incomes are present in the distribution.

Let us consider the case of extreme inequality when one family, say nth earns all the income and the other $(n-1)$ families together earn nothing at all. Now

$$k = n - 1, \quad y_n = 1, \quad \sum_{i=1}^{n-1} y_i = 0$$

and the expression (2.7) reduces to

$$\hat{G} = 1 + \frac{2}{n-1} \sum_{i=1}^{n-1} i \; y_i \qquad (2.8)$$

If all incomes were non-negative, then in the case of extreme inequality $y_i = 0$ for all $i < n$ and $\hat{G}$ equals one.

However, in the presence of negative income, we will have $y_i \gtreqless 0$ for $i \lessdot n$ and the term

$$\frac{2}{n-1} \sum_{i=1}^{n-1} i\, y_i \gtrdot 0 \qquad (2.9)$$

consequently $\hat{G}$ exceeds one.[3]

---

3. Since the extreme inequality is taken as a situation when $\sum_{i=1}^{n-1} y_i = 0$ and $y_n = 1$; if it is the unordered (equal weighted) case, $\sum_{i=1}^{n-1} y_i = 0$. Naturally, in the ordered case $\sum_{i=1}^{n-1} i\, y_i > 0$.

In general, $\hat{G} \gtrless 1$ depending on

$$\frac{1}{n-1} + \frac{2}{n-1}\left(\sum_{i=1}^{k} i\, y_i\right)$$

$$\gtrless \frac{1}{n-1} \sum_{i=k+1}^{n} y_i\,(1 + 2\,(n-1)) \tag{2.10}$$

As proposed by Chen *et al.* [2], a normalised Gini ratio $\hat{G}^*$ is defined as follows :

$$\hat{G}^* = \frac{1 + \frac{1}{n-1} + \frac{2}{n-1}\left(\sum_{i=1}^{n} i\, y_i\right) - \frac{1}{n-1} \sum_{i=k+1}^{n} y_i\,(1 + 2\,(n-1))}{\left(1 + \frac{2}{n-1} \sum_{i=1}^{k} i\, y_i\right)} \tag{2.11}$$

It is not difficult to realise that like the conventional Gini ratio $\hat{G}$, the normalised Gini ratio $\hat{G}^*$ lies between 0 and 1. For, in the case of equally distributed incomes $y_i = \frac{1}{n} \;\forall\; i$, and $\hat{G}^*$ reduces to $\hat{G}$, under (2.1), taking a value zero. And, in the event of extreme inequality,

$$\hat{G}^* = \frac{1 + \frac{1}{n-1} + \frac{2}{n-1} \sum_{i=1}^{n-1} i\, y_i - \frac{1}{n-1}}{1 + \frac{2}{n-1} \sum_{i=1}^{n-1} i\, y_i} \longrightarrow 1 \tag{2.12}$$

In the afore mentioned normalisation the definition of k (equation 2.6) becomes too special if one is handling empirical data. Normally, in most empirical situations we face data profiles, wherein there exists a value k', say, such that

$$\sum_{i=1}^{k'} y_i < 0 \quad \text{and} \quad \sum_{i=1}^{k'+1} y_i > 0 \tag{2.13}$$

There is thus a need to remodify $\hat{G}^{*}$ to get another version of normalised Gini ratio, namely $\hat{G}^{**}$.

Again, following Chen [2], Berrebi [1] and Chen [3], the entity

$$\frac{2}{n-1} \sum_{i=1}^{k'} i\, y_i \text{ is written as}$$

$$\frac{2}{n-1} \sum_{i=1}^{k'} i\, y_i + \frac{1}{n-1} \sum_{i=1}^{k'} y_i \left(\sum_{i=1}^{k'} \frac{y_i}{y_{k'+1}} - (1 + 2k') \right)$$

and the normalised Gini ratio takes the following form :

$$\hat{G}^{**} = \frac{A}{B}$$

where $A = 1 + \frac{1}{n-1} + \frac{2}{n-1} \sum_{i=1}^{k'} i\, y_i$

$$+ \frac{1}{n-1} \sum_{i=1}^{k'} y_i \left(- (1 + 2k') \right)$$

$$- \frac{1}{n-1} \sum_{i=k'+1}^{n} y_i \left[1 + 2\,(n - i) \right]$$

$$B = 1 + \frac{2}{n-1} \sum_{i=1}^{k'} i\, y_i + \frac{1}{(n-1)}$$

$$+ \sum_{i=1}^{k'} y_i \left\{ \sum_{i=1}^{k'} \frac{y_i}{y_{k'+1}} - (1 + 2k') \right\}$$

It can be easily seen, as in case of $\hat{G}^*$, that $\hat{G}^{**}$ also lies between 0 and 1.

Since the asymptotic distribution of $\hat{G}$ in the case of positive entries is available due to Ramakrishnan [8] and further since we have used in sequel its results and certain symbols and notations, we reproduce them in the form of Lemma that follows :

LEMMA : Asymptotic distribution of $\hat{G}$ is $N\left(G, \frac{\sigma_g^2}{n}\right)$ (2.14)

where $$\sigma_g^2 = \frac{\sigma_{11}}{4\mu^2} + \frac{G^2 \sigma_{22}}{\mu^2} - \frac{G\,\sigma_{12}}{\mu^2}$$

$$\sigma_{11} = \lim_{n\to\infty} n \operatorname{var}(\hat{\Delta}) = 4\, V[E \mid Y^* - Y_2 \mid],$$

and $V[E \mid Y^* - Y_2 \mid]$ = variance of the conditional expectation of $\mid Y_1 - Y_2 \mid$, given Y_1 is equal to some fixed value, say, Y^*,

$$\sigma_{12} = \lim_{n\to\infty} n \operatorname{cov}(\hat{\Delta}, \bar{Y}) = 2 \operatorname{Cov}[E \mid Y^* - Y_2 \mid, Y^*]$$

$$\sigma_{22} = \lim_{n\to\infty} n \operatorname{var}(\bar{Y}) = \sigma^2$$

G = Gini ratio for the population

$$= \frac{\Delta}{2\mu}$$

where, further, Δ = the population mean difference

μ = population mean.

Proof: The above results are proved by Ramkrishna [8], which are based on the theory of u-statistics and Hoeffding theorem [6]. It can be seen that $\hat{\Delta}$ and $\bar{X}$ are the corresponding u-statistics for estimable parameters Δ and μ.

Also observe that $2\sqrt{n}\,\bar{X}(\hat{G} - G) = \sqrt{n}(\hat{\Delta} - \Delta) - 2\,G\sqrt{n}(\bar{X} - \mu)$

and $\qquad \bar{X} \xrightarrow{p} \mu$

Invoking Slutsky's theorem (Serfling, [10], pp 19) and Hoeffding's result, it follows that

$$\sqrt{n}(\hat{G} - G) \xrightarrow{\text{asy}} N(0, \sigma_g^2)$$

Remark : In the above result the estimation of σ_g^2 is given by $\hat{\sigma}_g^2$ (Glasser, [4], pp 653) such that for large n and N (popn size), so that $\frac{n}{N} \lhd 1$,

$$\hat{\sigma}_g^2 = \frac{1}{\bar{Y}^2}\left[\text{var }\hat{\Delta}_i - 2\,\hat{G}\text{ cov }(\hat{\Delta}_i, Y_i) + \hat{G}^2\text{ var }Y_i\right] \tag{2.15}$$

where $\text{var }\hat{\Delta}_i = \frac{1}{n}\sum_{i=1}^{n}\hat{\Delta}_i^2 - (\hat{\Delta})^2$

$$\hat{\Delta}_i = \left[\frac{(2i - n - 1)\,Y_i + t_i}{n - 1}\right]$$

where, further, $t_i = \sum_{1 > i}^{n} Y_j - \sum_{j < i}^{n} Y_j$

$$\hat{\Delta} = \frac{1}{n}\sum_{i=1}^{n}\hat{\Delta}_i$$

$$\text{Cov }(\hat{\Delta}_i, Y_i) = \frac{1}{n}\sum_{i=1}^{n}\hat{\Delta}_i Y_i - \hat{\Delta}\,\bar{Y}$$

$$\text{Var }(Y_i) = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \bar{Y})^2$$

3. Asymptotic Distributions of $\hat{G}^*$ and $\hat{G}^{**}$

In what follows we obtain the sampling distributions of $\hat{G}^*$ and $\hat{G}^{**}$.

THEOREM 1. *$\hat{G}^*$ has the same limiting distribution as that of $\hat{G}$.*

Proof : The proof is based on the results of Slutsky's Theorem and the theorem on convergence in probability. Let us first recapitulate that

$$\frac{1}{n-1}\sum_{i=1}^{k} i\, y_i = \frac{2\sum\limits_{i=1}^{n} i\, Y_i}{n\,(n-1)\,Y} \tag{3.1}$$

$$E\left[\frac{2}{n\,(n-1)}\sum_{i=1}^{k} i\, Y_i\right] = \frac{\mu\, k^2}{n\,(n-1)} + \frac{\mu\, k}{n\,(n-1)}$$

$$\text{and} \quad \lim_{n\to\infty} E\left[\frac{2}{n\,(n-1)}\sum_{i=1}^{k} i\, Y_i\right] = 0 \tag{3.2}$$

The result (3.2) is obvious because $\frac{k}{n} \to 0$ as $n \to \infty$; since in any population (viable economy), the number of units possessing negative income is finite. Hence, by definition,

$$\frac{2}{n\,(n-1)}\sum_{i=1}^{k} i\, Y_i \xrightarrow{\text{mean}} 0$$

Since, convergence in mean implies convergence in probability, we have

$$\frac{2}{n\,(n-1)}\sum_{i=1}^{k} i\, Y_i \xrightarrow{p} 0$$

Hence, by theorem on convergence in probability,

$$2\sum_{i=1}^{k} i\, Y_i / n\,(n-1)\,\overline{Y} \xrightarrow{p} 0$$

which, implies that,

$$\frac{2}{n-1} \sum_{i=1}^{k} i\, y_i \xrightarrow{p} 0$$

$$\text{or } 1 + \frac{2}{n-1} \sum_{=1}^{k} i\, y_i \xrightarrow{p} 1 \tag{3.3}$$

By Slutsky's Theorem,

$$\frac{\hat{G}}{1 + \frac{2}{n-1} \sum_{i=1}^{k} i\, y_i} \text{ is asymptotically distributed as } N\left(G, \frac{\sigma_g^2}{n} \right).$$

Therefore, $\hat{G}^*$ has asymptotically the same distribution as that $\hat{G}$.

THEOREM 2. *Asymptotic distribution of* $\hat{G}^{**}$ *is same as that of* $\hat{G}$.

Proof: At the first instance, we will show that

$$\frac{1}{(n-1)} \sum_{i=1}^{k'} y_i \left(\sum_{i=1}^{k} \frac{y_i}{y_{k'+1}} - (1 + 2k') \right) \xrightarrow{p} 0$$

For this, we write its LHS as

$$\frac{1}{(n-1)} \frac{(\sum_{i=1}^{k'} y_i)^2}{y_{k'+1}} - \frac{(1 + 2k')}{(n-1)} \sum_{i=1}^{k'} y_i \tag{3.4}$$

we have

$$\sum_{i=1}^{k'} y_i = \frac{\sum_{i=1}^{k'} Y_i/n}{\bar{Y}}$$

Note that $\displaystyle\sum_{i=1}^{k'} \frac{Y_i}{n} \xrightarrow{p} 0$

and $\qquad \bar{Y} \xrightarrow{p} \mu$

Hence by theorem on convergence in probability, we have

$$\sum_{i=1}^{k'} y_i \xrightarrow{p} 0 \qquad (3.5)$$

or $\qquad (\sum_{i=1}^{k'} y_i)^2 \xrightarrow{p} 0$

Also observe that $(n-1)\, y_{k'+1} \xrightarrow{p} 1 \qquad (3,6)$

Combining (3.5) and (3.6) : and by theorem on convergence in probability, we have

$$\frac{(\sum_{i=1}^{k'} y_i)^2}{(n-1)\, y_{k'+1}} \xrightarrow{p} 0 \qquad (3.7)$$

It can easily be shown that

$$\frac{(1+2k')}{(n-1)} \sum_{i=1}^{k'} y_i \xrightarrow{p} 0 \qquad (3.8)$$

Again, invoking the theorem on convergence in probability and the results (3.7) and (3.8), we get

$$\left[\frac{(\sum_{i=1}^{k'} y_i)^2}{(n-1)\, y_{k+1}} - \frac{(1+2k')}{(n-1} \sum^{k'} y_i \right] \xrightarrow{p} 0$$

Having established this and using the result (3.3), we notice that the denominator in the expression for $\hat{G}^{**}$ converges to unity in probability,

$$\text{i.e. } 1 + \frac{2}{n-1} \sum_{i=1}^{k'} i\, y_i$$

$$+ \frac{1}{n-1} \sum_{i=1}^{k'} y_i \left\{ \frac{\sum_{i=1}^{k'} y_i}{y_{k'+1}} - (1+2k') \right\} \xrightarrow{p} 1 \tag{3.9}$$

Since $\hat{G}^{**}$ can also be written as

$$\hat{G}^{**} = \frac{\hat{G}}{1 + \frac{2}{n-1} \sum_{i=1}^{k'} i\, y_i + \frac{1}{(n-1)} y_i \left\{ \frac{\sum_{i=1}^{k'} y_i}{y_{k'+1}} - (1+2k') \right\}} \tag{3.10}$$

It immediately follows that the asymptotic distribution of $\hat{G}^{**}$ is same as that of $\hat{G}$. Thus

$$\hat{G}^{**} \sim N\left(G, \frac{\sigma_g^2}{n}\right) \tag{3.11}$$

Hence, whatever k we may encounter, the Gini ratio defined to include negative incomes as well, will have the same asymptotic distribution as that of conventional Gini ratio.

Once the asymptotic distributions of $\hat{G}^{*}$ and $\hat{G}^{**}$ are known, the related tests of significance based on single sample or two independent samples can easily be carried out for large n. For instance, let us take the case of two independent samples of sizes n_1 and n_2 respectively based on two distributions $F(y)$ and $G(y)$.

Let $\hat{G}_1^{*}$ and $\hat{G}_2^{*}$ be two estimates of Gini ratios based on these two samples.

In order to test whether these two coefficients differ significantly from each other, the appropriate test-statistic under $H_0 : \hat{G}_1^{*} = \hat{G}_2^{*}$, is given by

$$Z = \frac{\hat{G}_1^* - \hat{G}_2^*}{\sqrt{\frac{\hat{\sigma}_{g1}^2}{n_1} + \frac{\hat{\sigma}_{g2}^2}{n_2}}} \sim N(0, 1).$$

where

$$\frac{\hat{\sigma}_{g1}^2}{n_1} = \text{asymptotic variance of } \hat{G}_1.$$

$$\frac{\hat{\sigma}_{g2}^2}{n_2} = \text{asymptotic variance of } \hat{G}_2.$$

If we employ $\hat{G}^{**}$ instead of $\hat{G}^*$, the test-statistic will be

$$Z = (\hat{G}_1^{**} - \hat{G}_2^{**}) \Big/ \sqrt{\frac{\hat{\sigma}_{g1}^{*}}{n_1} + \frac{\hat{\sigma}_{g2}^{*}}{n_2}} \sim N(0,1).$$

REFERENCES

[1] Berrebi, Z. M. and Silber, J. (1985) : The Gini eoefficient and negative income, *Oxford Economic Papers*, 37 : 525-526.

[2] Chen-Tsauer-Rai (1982) : The Gini coefficient and negative income, *Oxford Economic Papers*, 34 : 473-478.

[3] Chen-Tsauer-Rai (1985) : The Gini coefficient and negative income : Reply, *Oxford Economie Papers*, 37 : 527-528.

[4] Glasser, G. J. (1962) : Variance formulas for the mean difference and coefficient of concentration, *JASA*, 57 : 648-54.

[5] Hagerbaumer, J. B. (1977) : The Gini concentration ratio and the minor concentration ratio : A two parameter index of inequality. *Review of Economics and Statistics*, 50 : 377-79.

[6] Hoeffding, W (1948) : A class of statistics with asymptotically normal distribution, *Ann. Math. Statist* 19 : 293-325.

[7] Pyatt, G., Chen, C. N. and Fei, J. C. H. (1980) : The distribution of income by factor components. *Quarterly Journal of Economics*, 95 : 451-73.

[8] Ramakrishanan, N. K. : Some results on Gini ratio. Discussion paper No. 1/84 (84), I.S.I., Bangalore.

[9] Sen, A. K. (1973) : *On Economic Inequality*. Oxford University Press, Oxford.

[10] Serfling, R. J. (1980) : *Approximation Theorems of Mathematical Statistics*. John Wiley and Sons, New York.

[11] Singh, Ajit (1971) : *Take overs, Their Relevance to the Stock Market and the Theory of the Firm*, Cambridge University Press, Cambridge.