REVIEW OF INEQUALITY METRICS WITH APPLICATION IN INCOME INEQUALITY IN LOUISIANA

Mihaela Paun, PhD[1]
Department of Mathematics and Statistics
College of Engineering and Science, Louisiana Tech University
P.O. Box 3178, Ruston, Louisiana 71272, U.S.A, E-mail: mpaun@latech.edu

Abstract

Economic data sets usually are, by their nature, very large and therefore researchers naturally want to analyze the distribution of the data set and make statistical inference about various parameters of interest, such as means, medians, variances, etc. To perform such tasks several incomes metrics or indices are generally used. Among the metrics widely used, we will consider in this paper only three: Lorenz curve, the the so-called S-Gini index and the Atkinson index. We also consider a general index of economic inequality that covers a number of indices, including the aforementioned S-Gini and Atkinson indices.

Keywords: Economic inequality, Gini index, Atkinson index, asymptotic normality.

1. Introduction

We are interested in analyzing the distribution of resources received by the population under consideration, in particular the disparity between the a percentage of population and the percentage of resources received. In particular, when one thinks about resources, one refers to income. Therefore we want to know what is the distribution of income among the individuals coming from the same population.

For performing such a task several metrics will be defined below. These metrics have values between 0 and 1. When a metric has value 0 we consider that the inequality is at minimum, which means each individual holds the same income. When the value of the metric is 1, inequality is considered to be at maximum, which means that one individual holds essentially all the income.

Knowing how to properly analyze this disparity inside a population and being able to compare several populations or the different distributions of the same population at different moments in time is of great importance for econometricians. Knowing how big the inequality among individuals is, one can derive taxation policies that can help the lower income individuals in the society, one can compensate for inflation or once the level of inequality is identified new measures can be implemented to reduce the inequality. Although there are several metrics that can be used for measuring inequality, there is no way of telling which one is the best measure, since the choice depends on the problem that we want to resolve or emphasize, or discover in the society under consideration.

In what follows we give main notions and notations used in subsequent sections. Specifically, we discuss some measures of variation among ``incomes'' such as the Gini coefficient, Lorenz curve and the Atkinson index. We also define a general index, discuss its asymptotic properties such as consistency and asymptotic normality, and then check those asymptotic properties in the special cases of the S-Gini and Atkinson indices.

2. Metrics of Economic Inequality

[*]Research was partially supported by Department of Mathematics and Statistics, Louisiana Tech University

Let X be a non-negative random variable, that models the income of individuals, with cumulative distribution function (cdf) $\boldsymbol{F}$.

For any $\boldsymbol{t} \in (0,1)$, the t × 100 percentile $\boldsymbol{F}^{-1}$ is given by the equation $\boldsymbol{F}^{-1}(\boldsymbol{t}) = \inf\{\boldsymbol{x} : \boldsymbol{F}(\boldsymbol{x}) \geq \boldsymbol{t}\}$, which defines the left-continuous inverse of the distribution function F.

Definition 1 *The* **Gini coefficient** *G is a measure of inequality in a population and is defined by the formula [3]:*

$$\boldsymbol{G} := \frac{1}{2\boldsymbol{\mu}}\mathrm{E}(|\boldsymbol{X}_1 - \boldsymbol{X}_2|) = \frac{1}{2\boldsymbol{\mu}}\int_{\boldsymbol{R}}\int_{\boldsymbol{R}} |\boldsymbol{x}_1 - \boldsymbol{x}_2| \, \boldsymbol{dF}(\boldsymbol{x}_1)\boldsymbol{dF}(\boldsymbol{x}_2), \tag{1}$$

where $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ are independent random variables.

The Gini coefficient was developed by the Italian statistician Corrado Gini in 1912 and a low Gini coefficient indicates more equal income or wealth distribution, while a high Gini coefficient indicates more unequal distribution.

We can derive from the formula above, the empirical Gini coefficient:

$$\begin{aligned}\boldsymbol{G}_{\boldsymbol{n}} : \quad &= \frac{1}{2\overline{\boldsymbol{X}}}\int_{\boldsymbol{R}}\int_{\boldsymbol{R}} |\boldsymbol{x} - \boldsymbol{y}| \, \boldsymbol{dF}_{\boldsymbol{n}}(\boldsymbol{x})\boldsymbol{dF}_{\boldsymbol{n}}(\boldsymbol{y}) \\ &= \frac{1}{2\boldsymbol{n}^2\overline{\boldsymbol{X}}}\sum_{\boldsymbol{i}=1}^{\boldsymbol{n}}\sum_{\boldsymbol{j}=1}^{\boldsymbol{n}} |\boldsymbol{X}_{\boldsymbol{i}} - \boldsymbol{X}_{\boldsymbol{j}}|,\end{aligned}$$

where

$$\boldsymbol{F}_{\boldsymbol{n}}(\boldsymbol{x}) = \frac{1}{\boldsymbol{n}}\sum_{\boldsymbol{i}=1}^{\boldsymbol{n}} \boldsymbol{I}(\boldsymbol{X}_{\boldsymbol{i}} \leq \boldsymbol{x}). \tag{2}$$

Definition 2 *The following metric is referred to, as the* ***Lorenz curve [4]***

$$L_F(z) := \frac{1}{\mu}\int_0^z F^{-1}(s)ds, \tag{3}$$

where $\boldsymbol{\mu}$ is the mean of $\boldsymbol{X}$ and $\boldsymbol{F}^{-1}(.)$ represents the left-continuous inverse of the distribution function F.

The Lorenz curve was developed by Max O. Lorenz in 1905 for representing income distribution. This is a graphical representation of the cumulative distribution function of a probability distribution, where the percentage of individuals is plotted on the x-axis and the percentage of income on the y-axis.

Definition 3 *The* ***Atkinson index****, denoted* $\boldsymbol{A}_{\boldsymbol{F}}$ *, is defined as follows [2]:*

$$A_F = 1 - \frac{1}{\mu}\left(\int_0^1 (F^{-1}(t))^a \, dt\right)^{\frac{1}{a}}, \tag{4}$$

where $\boldsymbol{a} > 0$ is a parameter, and $\boldsymbol{\mu}$ is the mean of $\boldsymbol{X}$.

In the above definition $\boldsymbol{\mu} = \boldsymbol{E}(\boldsymbol{X})$ denotes the mean of X which is assumed to be finite and non-zero,

$$0 < \boldsymbol{\mu} < \infty. \tag{5}$$

Recall that the Atkinson index $\boldsymbol{A_F}$ can be rewritten in the following form:

$$\boldsymbol{A_F} = 1 - \frac{1}{\boldsymbol{\mu}}(\mathbf{E}(\boldsymbol{X}^{\boldsymbol{a}}))^{\frac{1}{\boldsymbol{a}}}. \tag{6}$$

Using the corresponding empirical function for the quantile function defined in Eq. (2), we obtain the empirical Atkinson index:

$$A_n = 1 - \frac{1}{\overline{X}}\left(\int_0^1 (F_n^{-1}(t))^a \, dt\right)^{\frac{1}{a}}. \tag{7}$$

Studies dealing with the Gini index are extensive and the index is one of the principal inequality measure used in economics. However, in reality no explicit reason is given for preferring one measure of inequality over another. As such, we focus our attention next on the Atkinson index. The Atkinson Index is one of the few inequality metrics that incorporates normative judgments about social welfare [1][2]. The Atkinson index parameter a, is called the inequality aversion parameter. The parameter a reflects the strength of society's preference for equality, and can take values ranging from zero to infinity. When 0< a <1 the index suggests a preference for equality.

3. Louisiana Tech University Incomes - Case Study

In this case study, we present incomes from the Louisiana Tech University, for the year 2005-2006, [6]. We wish to perform a complete analysis on this data, using the indices defined above. We like to know how the salaries differ between 12 months group employess (administrative faculty, staff and maintanance) and the 9 months group employees (instructors, lecturers, tenure track or tenured faculty) and study where we have more inequality. We expect to find more inequality in the 12 months employees.

The first data set is a sample of n=558 salaries for the 12 months employees. A plot of these salaries is shown in Figure 1a. The minimum income in the sample is \$1,252 per year, the maximum is \$200,020 per year, and the mean salary is mean \$35,283.60 per year. To observe better how the incomes are spread in our data set, we present a histogram of the incomes in Figure 1b.

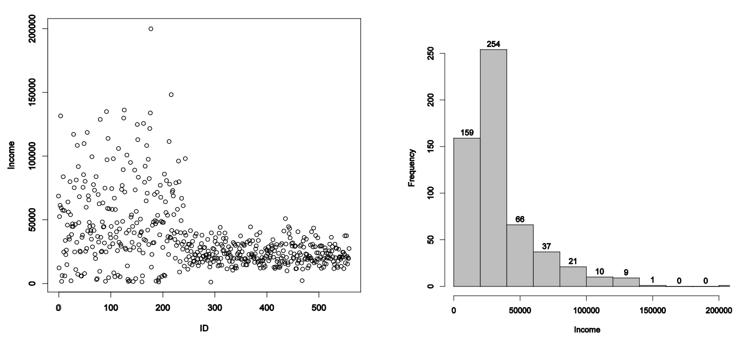

Figure 1. Plot and histogram of the yearly incomes for the 12 months employees at Louisiana Tech University

The second data set is a sample of $n = 336$ salaries for the 9 months employees. The minimum income in the sample is $\$4,990$ per year, the maximum is $\$129,289$ per year, and the mean salary is $ 51,452.94 per year.

Figures 2a and 2b are plots of the yearly incomes of the 9 months employees in the sample. To observe better how the incomes are spread in our data set, we present a histogram of the incomes.

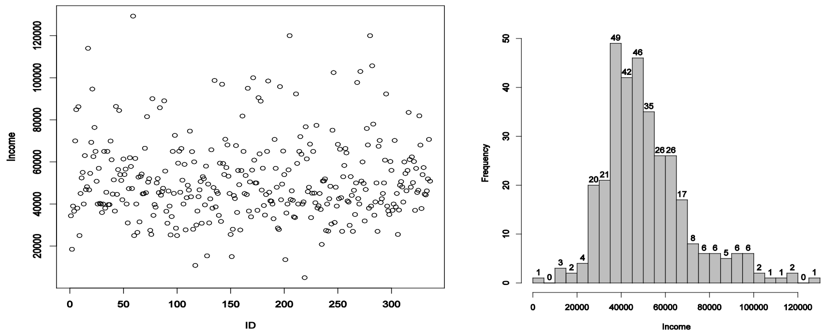

Figure 2. Plot and histogram of yearly incomes for the 9 months employees at Louisiana Tech University

If we just look at the means we clearly observe that 9 month employees have a larger salary average than the 12 month employees. However, the mean is not taking into consideration the fact that the data is skeewed or the possible outliers.

To have a better ideea about how big the inequality among these two data sets is, we show in Figure 3 two plots reprezenting the Lorenz curves for the two samples. The diagonal in the plots reprezents the exact equality in wealth distribution and the curve reprezents the distributional inequality. The greater the inequality, the more the line curves away from the diagonal.

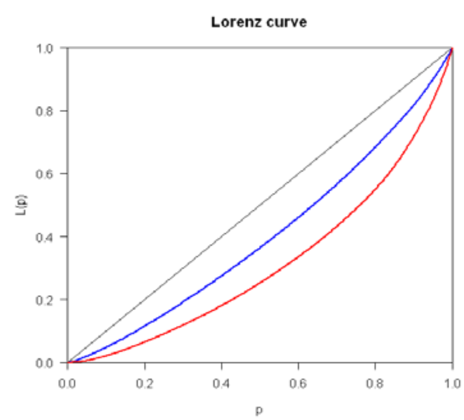

Figure 3. The Lorenz curves for the two data sets.

The curve closser to the diagonal (the red) reprezents the Lorenz curve for the 9 months employees and the curve further away from the diagonal (blue) reprezents the Lorenz curve for the 12 months employees. Since the Lorenz curve for the 9 months employees is closer to the diagonal, we conclude there is smaller inequality among the 9 months employees that the 12 months employees.

To analize the magnitude of the inequality and in which part of the distribution we have more inequality, we present Table 1.

Indices/Parameters	9 Months Employees	12 Months Employees
Gini Coefficient	0.19851	0.37272
Atkinson (a=0.1)	0.06528	0.02356
Atkinson (a=0.5)	0.03264	0.11477
Atkinson (a=1)	0.06592	0.22577
Atkinson (a=2)	0.14258	0.47204
Atkinson (a=10)	0.81492	0.93210

Table 1. Gini and Atkinson index for different values of the parameter.

Comparing the two Gini indices from Table 1, we observe that the inequality among the 12 months employees is three time biger than the inequality among the 9 months employees.

In Table 1 we calculated the Atkinson indices for different values of the parameter, for both samples. The small values of the Atkinson parameter emphasizes the high salaries and as the values of the parameter increase, the lower salaries are emphasized. We observe that although the 12 month salaries have a maximum greater than the 9 month salaries, there is more inequality among the top salaries in the 9 months data set (Atkinson = 0.06528) , than it is in the 12 months data set (Atkinson = 0.37272).

When the parameter is 1, we emphasize the middle size of the salaries distribution, and we observe that for the 9 month employees the middle size salaries are almost equal (Atkinson = 0.06592), compared to the 12 month employees salaries were we have some inequality (Atkinson = 0.22577). As the parameter increases and we get to emphasize the lower part of the distribution we observe that the level of inequality increeases (close to 1) for small salaries.

4. Conclusion

In line with our earlier assumption we find more inequality among the salaries of the 12 months employees compared to the inequality among the salaries of the 9 months employees. The analysis performed in Section 3 gives us a comprehensive image of how different these incomes are both within the same category and among the two categories.
These inequality metrics have the potential to be used in a large variety of application. Essentially every time we are interested in measuring any type of inequality we can apply these metrics. In [7], the Atkinson index was used for developing a framework of measuring performance and competence among employees. We are interested in finding new applications for these inequality metrics. In [6] we developed an asymptotic and bootstrap theory for the Atkinson index, for one and two populations. It is of interest to extend this theory when comparing inequality for more than two populations.

References

[1] Allison, Paul D., Measures of Inequality, *American Sociological Review*, 43, (1978), pp. 865-880.

[2] Atkinson, A. B., On the Measurement of Inequality, *Journal of Economic Theory* 2, (1970), 244-263.

[1] Gini, Corrado, "Measurement of Inequality and Incomes", *The Economic Journal* 31, (1921), 124-126.

[4] Lorenz, M. O., "Methods of measuring the concentration of wealth". *Publications of the American Statistical Association* 9, . (1905), 209-219.

[5] Gastwirth, Joseph L., "The Estimation of the Lorenz Curve and Gini Index". *The Review of Economics and Statistics* 54, (1972), 306-316.

[6] Păun, M. , Measuring Inequality: Statistical Inferential Theory with Applications, *Ph.D Disertation in Statistics, Louisiana Tech University, Louisnana*, May 2006.

[7] Păun, M. , Stanciu, M., "Competitiveness and performance - can we find a new measuring method?" *Review of Management and Economical Engineering*, special issue, vol.6, 2007, ISSN 1583-624X, pp.108-113.