

**A NOTE ON A SIMPLE INTERPRETATION OF THE GINI  
COEFFICIENT <sup>1</sup>  
WALTER SOSA ESCUDERO<sup>2</sup>**

Even though the Gini coefficient is likely the most frequently used measure of income inequality, policy makers and communicators rarely appreciate its conceptual and statistical advantages (see Cowell (2011) for an extensive discussion), and only exploit the fact that the higher its value the worse income inequality is in a certain region or period. Most intuitive presentations of the Gini coefficient are based on areas of the Lorenz curve, a rather abstract concept for many practitioners. This lack of obvious interpretability has led both experts and policy makers to cast doubts on the relevance of the Gini coefficient in practice. To the point, the widely cited book by Thomas Piketty (2014) claims that ‘Statistical indices such as the Gini coefficient give an abstract and sterile view of inequality...’. This short note provides an intuitive interpretation of the Gini coefficient based on simple averages of income differences, without resorting to the Lorenz curve explicitly, and hence being more amenable to those with minimal technical background.

Assume a society with  $n$  individuals with incomes  $y_1, y_2, \dots, y_n$ . Let  $\bar{y}$  be the average income and  $T$  the total income ( $T = \sum_{i=1}^n y_i = n \bar{y}$ ). Even though there are several equivalent ways to define the Gini index for this population (see Yitzhaki and Schechtman (2012) for a comprehensive review), a convenient one is the following (see Gasparini, Cicowiez and Sosa Escudero (2013) for a detailed discussion and empirical applications to the case of Latin America):

$$G = \frac{\sum_i \sum_j |y_i - y_j|}{n^2 2 \bar{y}}.$$

Let  $PD$  be the average of all differences among all income pairs, in absolute value:

$$PD = \frac{\sum_i \sum_j |y_i - y_j|}{n^2}.$$

---

<sup>1</sup> I thank Leonardo Gasparini for comments on an earlier version and two anonymous referees for helpful remarks that helped improve this note considerably.

<sup>2</sup> Universidad de San Andres and CONICET, Argentina. wsosa@udesa.edu.ar.

A convenient way to handle  $PD$  is through an inequality matrix  $M$ , with  $d_{ij} \equiv |y_i - y_j|$ :

$$M = \begin{vmatrix} d_{11} & d_{12} & \cdots & d_{1n} \\ d_{21} & d_{22} & & \\ \vdots & & \ddots & \vdots \\ d_{n1} & \cdots & & d_{nn} \end{vmatrix}.$$

Clearly,  $PD$  is the average of the  $n^2$  elements of  $M$ . Intuitively, the Gini coefficient measures how ‘large’ is this matrix of income disparities, in a relative sense. In a completely egalitarian society, all individuals have the same income ( $y_i = y_j = \bar{y}$ ), hence all elements of  $M$  are zero. In the case of extreme inequality one individual has the whole income and the rest nothing. Assuming (without loss of generality) that the only individual that has everything is the last one. The  $M$  matrix for this case ( $M_T$ ) is:

$$M_T = \begin{vmatrix} 0 & 0 & \cdots & 0 & T \\ 0 & 0 & & 0 & T \\ \vdots & & \ddots & & \vdots \\ 0 & 0 & \cdots & 0 & T \\ T & T & \cdots & T & 0 \end{vmatrix}.$$

Naturally,  $PD$  is zero for the completely egalitarian case. It is easy to see that  $PD$  for the other extreme case ( $PD_T$ ) is:

$$PD_T = \frac{2(n-1)T}{n^2} = 2 \frac{n-1}{n} \frac{T}{n} \sim 2\bar{y},$$

where the last expression is valid for large  $n$ . Then,

$$G = \frac{PD}{PD_T}.$$

For example, if in a society of three individuals incomes in dollars are 1,2 and 9,  $T = 12$  and  $\bar{y} = 4$ . The average of all income discrepancies ( $PD$ ) is 3.55. If all income is concentrated in only one person  $PD_T$  is 5.33 (using the exact formula, since in this case  $n = 3$  is very small). Then, the Gini coefficient for this example is  $G = 3.55/5.33 = 0.66$ .

The previous expression shows that  $G = 0$  if the distribution of incomes is completely egalitarian ( $PD = 0$ ) and  $G = 1$  in the other extreme case ( $PD = PD_T$ ). Obviously,  $G \geq 0$  since  $PD \geq 0$ , by construction. If all incomes are non-negative, based on the triangle inequality it is easy to show (see the Appendix) that  $PD \leq PD_T$ , hence  $0 \leq G \leq 1$ .

To summarize, the Gini coefficient measures how far are the aggregate income discrepancies of a society with respect to two extreme societies: one completely egalitarian ( $G = 0$ ) and another one with maximal inequality ( $G = 1$ ). A final remark is that computing the Gini coefficient using average discrepancies has no obvious empirical advantages with respect to alternative formulae based the Lorenz curve, since all of them led to consistent estimators of the population Gini index.

## References

- Cowell, F. (2011). “Measuring Inequality,” Oxford University Press, Oxford.
- Gasparini, L., Cicowiez, M. y Sosa Escudero, W. (2013). “Pobreza y Desigualdad en América Latina. Conceptos y Herramientas Analíticas,” Editorial Temas, Buenos Aires.
- Piketty, T. (2014). “Capital in the Twenty-First Century,” Harvard University Press, Cambridge.
- Yitzhaki, S. y Schechtman, E. (2012). “More Than a Dozen Alternative Ways of Spelling Gini,” chapter 2 in Yitzhaki y Schechtman, *The Gini Methodology*, Springer, New York.

## Appendix: Proof that $PD \leq PD_T$ .

By the triangle inequality

$$|y_i - y_j| \leq |y_i| + |-y_j|$$

$$y_i + y_j,$$

since all incomes are non-negative. Then

$$\frac{1}{n^2} \sum_i \sum_j |y_i - y_j| \leq \frac{1}{n^2} \sum_i \sum_j (y_i + y_j) = 2\bar{y}$$

hence  $PD \leq PD_T$ .