Economics Letters
Manuscript Draft

Manuscript Number: EL29122

Title: A Note on the Relationship between Top Income Shares and the Gini Coefficient

Article Type: Original Article

Keywords: Gini coefficient; top incomes.

Abstract: When a very top group of the income distribution, infinitesimal in numbers, owns a finite share S of total income, then the Gini coefficient G can be approximated by G*(1-S)+S, where G* is the Gini coefficient for the rest of the population. We provide a simple formal proof for this expression and offer two applications as illustrations.

A Note on the Relationship between Top Income Shares and the Gini Coefficient

Facundo Alvaredo*
Oxford University, CONICET and CEPR

January 2010

Abstract

When a very top group of the income distribution, infinitesimal in numbers, owns a finite share S of total income, then the Gini coefficient G can be approximated by $G^*(1-S)+S$, where G^* is the Gini coefficient for the rest of the population. We provide a simple formal proof for this expression and offer two applications as illustrations.
JEL D31 H2

1 Introduction

In a typical income distribution, the rich may appear insignificant. The most commonly used measure of inequality, the Gini coefficient, is more sensitive to transfers at the center of the distribution that at the tails. In a textbook-sized Lorenz curve, the top 0.1% or even the top 1% are scarcely distinguishable on the horizontal axis from the vertical endpoint. However, changes in top income shares are capable of impacting on changes in overall inequality significantly, as advanced by Atkinson (2007): "If we treat the very top group as infinitesimal in numbers, but with a finite share S of total income, then the Gini coefficient G can be approximated by $G^*(1-S)+S$, where G^* is the Gini coefficient for the rest of the population" (p. 19). The relevance of the last expression has increased with the recent developments of the literature on top incomes (Atkinson and Piketty, 2007, 2010) and the comparison of inequality statistics from survey data and tax records (Burkhauser et al., 2009).

*Email: facundo.alvaredo@economics.ox.ac.uk. Address: Department of Economics, Oxford University, Manor Road Building, Manor Road, OX1 3UQ Oxford UK. I thank Tony Atkinson, Maria Ana Lugo and Thomas Piketty for comments and discussions.

The purpose of this note is to provide a simple formal proof of the last statement about the connection between top income shares and the Gini coefficient (not given in Atkinson, 2007) and to offer some illustrative examples of its application.

There are several ways to arrive to the same conclusion. From a graphical perspective, the result is rather intuitive: when the very top group owns a large share of total income S, the Lorenz curve $L(p)$ almost touches the right y-axis at $1 - S$. Let's call $L^*(p)$ the Lorenz curve for the non-top population (the bottom 99%, the bottom 99.9%, etc.). Given that, for the non-top group, $L(p) \simeq L^*(p)(1 - S)$, and that the Gini coefficient G (in continuous space) is $1 - 2\int L(p)dp$, then it is straightforward to note that $G \simeq 1 - 2\int L^*(p)(1 - S)dp \simeq G^*(1 - S) + S$. More formally, we start from the decomposition of the Gini coefficient in discrete space proposed by Dagum (1997).

2 The Decomposition of the Gini Coefficient

Let's consider a population on N individuals with mean income μ, partitioned in $j = 1, 2, ..., k$ subpopulations of N_j individuals with mean income μ_j. Each individual i in subgroup j has income y_{ij}. The Gini coefficient of the whole population is

$$G = \frac{\sum_{j=1}^{k} \sum_{h=1}^{k} \sum_{i=1}^{N_j} \sum_{r=1}^{N_h} |y_{ij} - y_{hr}|}{2N^2\mu} \quad (1)$$

The Gini coefficient *within* the j-th subpopulation (simply the Gini of the j-th subpopulation) is

$$G_{jj} = \frac{\sum_{i=1}^{N_j} \sum_{r=1}^{N_j} |y_{ji} - y_{jr}|}{2N_j^2\mu_j} \quad (2)$$

The Gini coefficient *between* the j-th and the h-th subpopulations is (Dagum, 1987)

$$G_{jh} = \frac{\sum_{i=1}^{N_j} \sum_{r=1}^{N_h} |y_{ij} - y_{hr}|}{N_j N_h(\mu_j + \mu_h)} \quad (3)$$

from which it is straightforward to note that $G_{jh} = G_{hj}$.

Let P_j be the j-th subpopulation share in total population

$$P_j = \frac{N_j}{N}$$

and S_j the j-th subpopulation income share

$$S_j = \frac{N_j \mu_j}{N \mu}$$

Dagum (1997) has shown that the Gini coefficient for the whole population can be decomposed as follows:

$$G = \sum_{j=1}^{k} G_{jj} P_j S_j + \sum_{j=1}^{k} \sum_{h=1}^{j-1} G_{jh}(P_j S_h + P_h S_j) = G_w + G_b \tag{4}$$

G_w measures the contribution of inequality *within* subpopulations, and G_b measures the contribution of inequality *between* subpopulations.

3 Top Income Shares

We consider a population partitioned in two (k=2). In subgroup $j = 1$ we have individuals at the top of the distribution (e.g. the top 0.01%, the top 0.1%, etc.), with income share S and population share P. The rest of the population is in subgroup $j = 2$, with income share $1 - S$ and population share $1 - P$. Then (4) can be expressed as

$$\begin{aligned} G &= G_{11}PS + G_{22}(1-P)(1-S) + G_{12}P(1-S) + G_{21}(1-P)S \\ &= \underbrace{G_{11}PS + G_{22}(1-P)(1-S)}_{G_w} + \underbrace{G_{12}(P(1-S) + (1-P)S)}_{G_b} \end{aligned} \tag{5}$$

In this case (with only 2 groups and with higher-income individuals in j=1), G_b in equation 5 can be further simplified:

$$\begin{aligned} G_b &= G_{12}(P(1-S) + (1-P)S) \\ &= \frac{\sum_{i=1}^{N_1}\sum_{r=1}^{N_2}(y_{1i} - y_{2r})}{N_1 N_2(\mu_1 + \mu_2)}(P(1-S) + (1-P)S) \\ &= \frac{\mu_1 - \mu_2}{\mu_1 + \mu_2}(P(1-S) + (1-P)S) \\ &= \frac{\mu_1 - \mu_2}{\mu_1 + \mu_2}P(1-P)\frac{\mu_1 + \mu_2}{\mu} \\ &= (1-P)S - P(1-S) \\ &= S - P \end{aligned} \tag{6}$$

This is equivalent to the result described graphically in Atkinson and Bourguignon (2000), pp. 7-8, for the two-class case. Incorporating (6) in (5) we get

$$G = G_{11}PS + G_{22}(1-P)(1-S) + S - P \tag{7}$$

For a very top subgroup, infinitesimal in numbers ($P \to 0$), but with a finite share S of total income, we have

$$\lim_{P \to 0}[G_{11}PS + G_{22}(1-P)(1-S) + S - P] = G_{22}(1-S) + S \tag{8}$$

In expression (8), G_{22} is the Gini coefficient of the non-top subpopulation, called G^* in the introduction.

4 Applications

4.1 Case 1: United States

Burkhauser et al. (2009) have tried to reconcile Piketty and Saez (2003) tax-based top income share series with top income shares from the United States internal CPS. The internal CPS is less affected by top code than the public CPS. They find that their CPS-based top income shares series closely match the Piketty and Saez (2003) series for the top 10-1% (the top decile excluding the top percentile). However, even if the top-code effect is less pervasive, the top 1% measured by the internal CPS is consistently lower than the top 1% measured with tax data.

As described in Atkinson, Piketty and Saez (2009), the official CPS Gini in the US increased from 39.8 in 1976 to 47.0 in 2006, the change between those two years (net of measurement adjustments in 1992-1993) being 5.3 percentage points. Using the formula in (8) and the top 1% share from Burkhauser et al. (2009), the authors estimated G^* for the bottom 99% of the population (results reproduced in Table 1) and then computed G using G^* and the top 1% share from tax data. It turned out that G increased from 41.1 to 51.9 (top share including capital gains) and from 40.5 to 49.3 (top share excluding capital gains) over the same period. If the series including capital gains are taken as benchmark, then the rise in G, 10.8 percentage points, is more than twice as large as the 5.3 percentage points increase recorded by the official series, and more than three times as large as the 3.2 percentage points rise in G^*.[1] As the authors state, "the top percentile plays a major role in the increase in the Gini over the last three decades and CPS data which do not measure top incomes fail to capture about half of this increase in overall inequality."

[1]These results should be taken as approximations, as top income share estimates from Burkhauser et al. (2009) refer to the family distribution excluding cash transfers, the official CPS Gini is based on households income including cash transfers, and the unit of analysis in Piketty and Saez (2003) is the individual.

4.2 Case 2: Argentina

Székeley and Hilgert (1999) have analyzed a large number of Latin American surveys to confirm that surveys' top incomes generally correspond to the prototype of highly educated professionals rather than capital owners. They find that the income of the ten richest households in the surveys is generally similar to the average wage of a manager of a medium to large size firm (and, in many cases, even below that level). This observation has important implications, as survey-based results have pointed to a decline in inequality in Latin America over the last years (Lopez-Calva and Lustig, 2010).

We take the case of Argentina as the second illustrative example. Table 2 displays the tax-based top 1% and top 0.1% income shares from Alvaredo (2010) and the survey-based Gini coefficient (G^*) between 1997 and 2004. Using the formula in (8), we computed G in two hypothetical cases, namely that the top 1% and the top 0.1% are not represented in the surveys.[2] Results are shown in columns 4 and 5, respectively. Two unsurprising facts are readily noticeable. Firstly, G can be several percentage points above G^*. Secondly, not only can levels be different, but also the trends of G and G^* can diverge. According to the survey's results, G^* displays virtually no change when 2001 and 2003 are compared, going from 51.1 to 50.9. However, G "corrected" with the top 1% income share (column 5) was 57.4 in 2001 and 59.2 in 2003 (almost a two percentage points increase).

If top incomes ignored by surveys experience a large enough relative increase, then the true dynamics of overall inequality may display a rising trend even when survey-based estimates show opposite results. As long as surveys do not record what is happening with the true distribution at the top, survey-based estimates showing a decline in inequality can *at most* indicate that those reductions are happening within non-top individuals.

5 References

Alvaredo, F. (2010). The Rich in Argentina over the Twentieth Century, in A. Atkinson and T. Piketty, op. cit.

[2]Slightly different from the strategy followed in the previous subsection, in this case we assume that top individuals are completely ignored by the survey, so we directly consider G^* the result arising from the whole sample.

Atkinson, A. B. (2007). Measuring Top Incomes: Methodological Issues, in A. Atkinson and T. Piketty, op. cit.

Atkinson, A. B. and Bourguignon, F. (2000). Introduction: Income Distribution and Economics, in A. B. Atkinson and F. Bourguignon (editors), *Handbook of Income Distribution*, vol i, Elsevier.

Atkinson, A. B. and Piketty, T. (2007). *Top Incomes over the Twentieth Century: A Contrast Between Continental European and English-Speaking Countries*, Oxford: Oxford University Press.

Atkinson, A. B. and Piketty, T. (2010). *Top Incomes over the Twentieth Century: A Global View*, Oxford: Oxford University Press.

Atkinson, A. B., Piketty, T. and Saez, E. (2009). Top Incomes in the Long Run of History. NBER WP 15408.

Burkhauser, R., Feng, S., Jenkins, S. and Larrimore, J. (2009). Recent Trends in Top Income Shares in the USA: Reconciling Estimates from March CPS and IRS Tax Return Data, NBER WP 15320.

Dagum, C. (1987). Measuring the Economic Affluence between Populations of Income Receivers, *Journal of Business and Economic Studies*, 5(1): 5-12.

Dagum, C. (1997). A New Approach to the Decomposition of the Gini Income Inequality Ratio, *Empirical Economics*, 22: 515-531.

López-Calva, F. and Lustig, N. (2010). *Declining Inequality in Latin America: A Decade of Progress?*, Washington DC: The Brookings Institution.

Székeley, M. and Hilgert, M. (1999). What's behind the inequality we measure: and investigation using Latin American data. Research Department WP 409. Inter-American Development Bank.

Table 1

Table 1. Top income shares and the Gini coefficient in the US, 1976 and 2006

	top 1% income share CPS data	top 1% income share tax data excluding K gains	top 1% income share tax data including K gains	Gini Coeff. CPS data	Gini Coeff. G* bottom 99%	Gini Coeff. G corrected with tax-based top 1% excluding K	Gini Coeff. G corrected with tax-based top 1% including K gains
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
1976	6.7	7.9	8.9	39.8	35.5	40.5	41.1
2006	13.7	18.0	22.8	47.0	38.6	49.3	51.9
Point increase	6.9	10.1	14.0	7.2	3.2	8.8	10.8
Point Increase removing the '92-'93 CPS discontinuity	4.1			5.3	3.2		

Sources: Results reproduced from Atkinson, Piketty and Saez (2009).

Notes: Following expression (8), and using the top 1% income share from CPS data (column 1) and the official Gini from CPS data (column 4), G* for the bottom 99% for 1976 in column 5 is computed as 100*(0.398-0.067)/(1-0.067). Analogously, using the tax-based top 1% income share including capital gains (column 3) and G* (column 5), the Gini coefficient G, corrected with the tax-based top 1% income share including capital gains for 1976 in column 7 is computed as 100*(0.355*(1-0.089)+0.089).

Discrepancies are due to rounding.

Table 2

Table 2. Top income shares and the Gini coefficient in Argentina, 1997-2004

	top 1% income share (%) (1)	top 0.1% income share (%) (2)	Gini Coeff. G* (3)	Gini Coeff. G corrected with the top 0.1% income sh≀ (4)	Gini Coeff. G corrected with the ≀op 1% income shar (5)
1997	12.39	4.27	46.9	49.2	53.5
1998	12.57	4.37	48.5	50.8	55.0
1999	13.53	5.22	47.0	49.7	54.1
2000	14.34	5.68	48.6	51.6	56.0
2001	12.91	5.22	51.1	53.6	57.4
2002	15.53	6.92	51.9	55.2	59.4
2003	16.85	7.40	50.9	54.6	59.2
2004	16.75	7.02	48.8	52.4	57.3

Notes: Top shares in columns (1) and (2) are taken from Alvaredo (2010).
G* denotes the Gini coefficient of individual income based on the Greater Buenos Aires
household survey. All results correspond to
October surveys, except for 2003 (May). Only income earners with positive income were considered and no
further adjustments were applied. The Greater Buenos Aires households survey is taken as representative of Argentina.
Following expression (8), G is computed as S+(1-S)G*, where S is the estimate of the top 0.1% income share in the case of column 4,
and the estimate of the top 1% income share in the case of column 5.