Mathematics Magazine

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/umma20

Income Inequality and Taxation

Robert A. Agnew

To cite this article: Robert A. Agnew (2021) Income Inequality and Taxation, Mathematics Magazine, 94:4, 257-266, DOI: 10.1080/0025570X.2021.1957393

To link to this article: https://doi.org/10.1080/0025570X.2021.1957393

Published online: 10 Sep 2021.

Income Inequality and Taxation

ROBERT A. AGNEW
Palm Coast, FL 32137
raagnew1@gmail.com

It is hard to pick up a magazine these days without encountering an extensive article on income inequality, wealth inequality or both. Moreover, acclaimed books by economists Piketty [9], Reich [10], Stiglitz [13], and most recently Saez and Zuchman [11] have highlighted the issue for the general public, while Democratic presidential candidates have highlighted comparisons between the top 1% and the bottom 99%. There is no doubt that income inequality is a serious economic and political issue, particularly in the United States. The purpose of this article is to indicate, using straightforward mathematics, how income inequality can be reduced significantly via a simple linear income tax and to illustrate this with data from the United States Census Bureau and Internal Revenue Service (IRS).

Fontenot, Semega, and Kollar [4], in Table 2, present two traditional measures of income inequality:

(1) Shares of aggregate household income received by population quintiles (and the top 5%).

(2) The Gini index, plus the Theil index (similar to Gini) and other auxiliary metrics.

The Gini index attempts to capture distributional income inequality in one single number. Some, including Piketty, take issue with this view, but we focus entirely on that particular measure.

In 2017, per Census, 22.3% of the U.S. household income went to the top 5% of income-earners, while only 25.6% went to the bottom 60% of income-earners. The U.S. Gini index was 0.482, which Stiglitz notes is high relative to other developed nations and significantly increased from 1980. The IRS does not focus on inequality, but analysis of their tabulated data indicates that our current complex tax system does little to alleviate it. Indeed, average tax rates are not progressively higher at the top end of the adjusted gross income scale. Once again, these figures highlight the extent of the problem in the U.S.

Mathematicians were recently introduced to the Gini index, and various approximations to it, by Farris [3] and Jantzen and Volpert [6]. These authors have demonstrated that there is some interesting mathematics associated with income inequality measurement. Moreover, Farris developed a useful interpolation scheme for stratified income tables that we utilize later.

The next section reviews the Gini index, both pre-tax and post-tax, in general terms. We then get more specific with the linear income tax, much studied by economists [1, 8, 12], as well as in more recent economic reviews of optimal taxation [2, 7]. We derive the relationship between pre-tax and post-tax Gini, which indicates that significant inequality reduction is possible with a very simple linear tax structure. In the final section, we benchmark against actual U.S. income data from the Census Bureau [14] and the IRS [15].

Math. Mag. **94** (2021) 257–266. doi:10.1080/0025570X.2021.1957393 © Mathematical Association of America
MSC: 91B15

Income inequality measurement: The Gini index

We start with the Lorenz curve $L(p)$ which defines the cumulative share of total money income associated with the lowest-income proportion p of a population. In other words, the cumulative fraction p of the population gets cumulative fraction $L(p)$ of total money income, arranged in order of increasing money income rate. To illustrate this concept, we already saw from 2017 Census data that $L(.6) = 0.256$ and $L(.95) = 1 - .223 = 0.777$. Mathematically, the Lorenz curve is a nondecreasing, convex function from [0, 1] onto itself with $L(0) = 0$ and $L(1) = 1$, while we normally expect the curve to be strictly increasing and strictly convex throughout. In this context, $L'(p)$ can be viewed as either a probability density function on the unit interval, or alternatively as the scaled money income rate at the lowest-income proportion p of the population. We will take the latter view. We assume throughout that we can treat L as twice continuously differentiable on the unit interval, notwithstanding the inherent discreteness in any real population.

From Farris [3], the Gini index of income inequality is defined by

$$G = 2 \int_0^1 (p - L(p))\, dp \tag{1}$$

The Gini index is portrayed graphically in Figure 1 as twice the area between the equity reference line and the Lorenz curve.

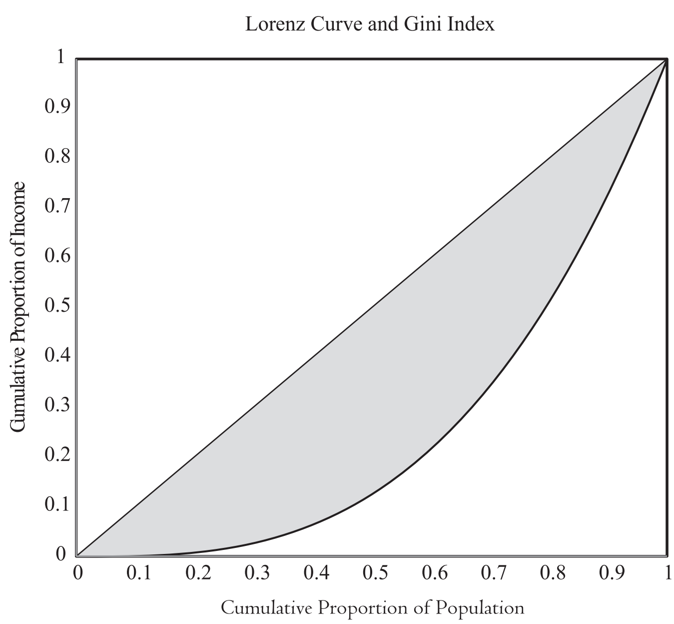

Figure 1 The Gini index is twice the area between equity reference line and the Lorenz curve.

A Gini of zero indicates perfect equality (per the equity reference line), whereas a Gini of one indicates perfect inequality in a limiting sense (per the boundary lines). As indicated previously, various approximations to both the Lorenz curve and the associated Gini index have been suggested, but for our derivation we simply assume that the Lorenz curve is given and we do not worry about its exact functional form.

We now introduce taxation as follows: Relabel the pre-tax Lorenz curve as L_0 and its associated Gini index as G_0. We introduce the cumulative tax curve as $F(p)$, the cumulative proportion of pre-tax money income received as tax from the lowest-income proportion p of the population. Then $F(0) = 0$ and $F(1) = c$, where $c \in (0, 1)$ represents the total proportion of pre-tax money income collected by the government. In this setup, $F'(p)$ is the scaled monetary tax rate at the lowest-income proportion p and $F'(p)/L_0'(p)$ is the percentage tax rate at p. For simplicity, our focus is entirely on personal or individual income taxation at the federal level. That is, we ignore corporate income taxes, all state and local taxes, and all other federal taxes.

We expect F' to be increasing (i.e., $F'' > 0$) throughout, but we need $F'(p)/L_0'(p)$ to be increasing throughout for a progressive percentage tax rate schedule. We can certainly have $F(p) < 0$ and $F'(p) < 0$ for lower portions of the curve, corresponding to a negative income tax which subsidizes lower-income individuals and is fundamental to significantly reducing inequality. It is worth noting that the negative income tax was favored by Milton Friedman [**5**] as a means of alleviating poverty on the lower end of the income scale.

We now define $H(p) = L_0(p) - F(p)$ as cumulative post-tax income as a proportion of total pre-tax income. We understand that the government is a large-scale employer, but there is no inconsistency if pre-tax income of government workers is included in L_0. We then have $H(0) = 0$, $H(1) = 1 - c$, and we must have both $H' > 0$ and $H'' > 0$ to make sense. The basic curves are illustrated schematically in Figure 2.

We now define the post-tax Lorenz curve as

$$L_1(p) = \frac{L_0(p) - F(p)}{1 - c}$$

so that $L_1(0) = 0$ and $L_1(1) = 1$. Then the post-tax Gini index is defined by

$$\begin{aligned} G_1 &= 2\int_0^1 (p - L_1(p))\,dp \\ &= 2\int_0^1 \left(p - \frac{L_0(p) - F(p)}{1 - c}\right) dp \\ &= G_0 - \frac{2\int_0^1 (cL_0(p) - F(p))\,dp}{1 - c} \end{aligned} \tag{2}$$

so that Gini reduction is proportional to the area between two cumulative curves, $cL_0(p)$ and $F(p)$, both joining points $(0, 0)$ and $(1, c)$ (see Figure 2). The curve $cL_0(p)$ is c times the solid line (plotted as dash-dot), and $F(p)$ is the dotted line. The curve $cL_0(p)$ is the cumulative tax curve associated with a non-progressive "flat" tax with constant tax rate $F'(p)/L_0'(p) = c$ across the whole income spectrum. Hence, from (2), $G_1 = G_0$ for a flat tax and there is no Gini reduction. We will investigate this relationship further in the next section for a full-fledged linear income tax. Figure 3 depicts a progressive tax schedule yielding significant Gini reduction.

Linear income tax

The linear income tax is defined by

$$F'(p) = rL_0'(p) - s$$

where $r \in (0, 1)$ is the marginal percentage tax rate and s is a common scaled monetary subsidy that everybody gets, often called "universal basic income." In this context, the

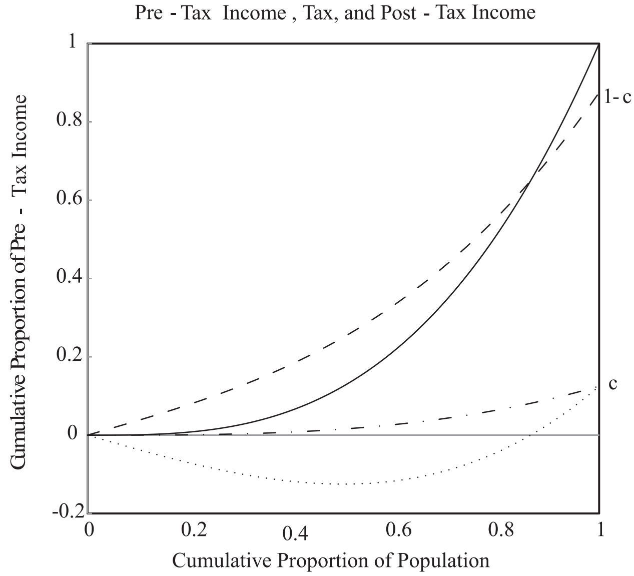

Figure 2 The solid line is pre tax income. Tax is dotted, post-tax income is dashed, and the flat-tax dash-dot, all proportional to pre-tax income. The tax curve can be negative on the low end. The Gini reduction is proportional to the area between the flat tax and tax curves.

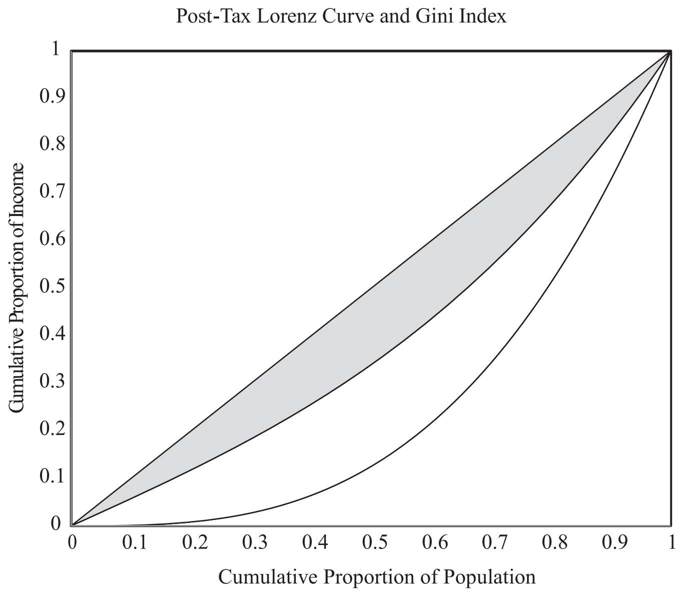

Figure 3 Post-tax Gini index is twice the area between the equity reference line and the post-tax Lorenz curve. The pre-tax Lorenz curve is also shown.

tax schedule is progressive if $F'(p)/L_0'(p)$ is increasing, which is true if $s > 0$. Upon integrating, we get $F(p) = rL_0(p) - sp$. Since $F(1) = c$, we have that $c = r - s$ or $s = r - c$. Hence, the cumulative tax curve is

$$F(p) = rL_0(p) - (r - c)p \tag{3}$$

and progressivity requires that $r > c$. Moreover, if

$$rL_0'(0) < r - c$$

we will have negative taxation on the lower end of the income scale. If $r = c$, we have the non-progressive flat tax referenced in the last section. If $r < c$, we have a regressive tax schedule with negative subsidy.

The general linear income tax has been studied extensively by economists in the context of both fair distribution and economic efficiency [**1**, **8**, **12**]. In particular, Aumann and Kurz [**1**] derived it via the Shapley value of cooperative game theory in the context of a majority–minority political game with threats. Coming from somewhat different directions, all these economists arrive at a version of (3) with positive subsidies and marginal tax rates in the 50%+ range, depending on utility assumptions. For linear monetary utility, Aumann and Kurz get exactly $r = 50\%$ with higher marginal tax rates for more concave monetary utility functions. Now we turn to calculation of post-tax Gini for the general linear income tax. We prove the following proposition:

Proposition. *For pre-tax Lorenz curve L_0, associated Gini index G_0, and cumulative tax curve defined by (3), the post-tax Gini index is*

$$G_1 = \left(\frac{1-r}{1-c}\right) G_0 \tag{4}$$

with percentage Gini reduction

$$1 - \frac{G_1}{G_0} = \frac{r-c}{1-c} \tag{5}$$

Proof. Using (3), we have

$$\begin{aligned} 2\int_0^1 (cL_0(p) - F(p))\,dp &= 2\int_0^1 (cL_0(p) - rL_0(p) + (r-c)p)\,dp \\ &= (r-c)G_0. \end{aligned}$$

Hence from (2),

$$G_1 = G_0 - \left(\frac{r-c}{1-c}\right) G_0 = \left(\frac{1-r}{1-c}\right) G_0,$$

and further,

$$1 - \frac{G_1}{G_0} = \frac{r-c}{1-c}.$$

■

Under a linear income tax structure, we conclude that inequality reduction is a function of the relationship between r and c. If r is much larger than c, then significant inequality reduction is possible. In the next section, we refer to empirical data from the U.S. Census Bureau and the U.S. Internal Revenue Service.

Benchmarking against the U.S. income data

So far our development has been continuous and fairly abstract. We now want to benchmark against actual tabulated U.S. income data, from both the Census Bureau and the IRS. In the interest of space, we will not reproduce their tables here. Instead, we will simply present highlights. To begin, we adopt the tabular notation of Farris [3] where we have n household strata arranged from lowest to highest average income with the following definitions:

$$h_i = \text{Number of Households in Stratum } i$$

$$N = \sum_{i=1}^{n} h_i = \text{Total Number of Households}$$

$$p_j = \sum_{i=1}^{j} h_i/N = \text{Cumulative Household Proportion through Stratum } j$$

$$x_i = \text{Average U.S. Dollar Annual Pre-Tax Money Income per Household in Stratum } i$$

$$T = \sum_{i=1}^{n} x_i h_i = \text{Total Household Dollar Income}$$

$$\bar{x} = T/N = \text{Overall Average Dollar Income}$$

From these ingredients, Farris noted that the pre-tax Lorenz curve

$$L_0(p_j) = \sum_{i=1}^{j} \frac{x_i h_i}{T}$$

through stratum j, and then he derived the following interpolation formula for the pre-tax Gini index:

$$G_0 \cong \left(\frac{1}{T} \sum_{j=1}^{n} x_j (p_j + p_{j-1}) h_j \right) - 1 \tag{6}$$

where $p_0 = 0$. An equivalent interpolation formula is

$$\begin{aligned} G_0 &\cong 1 - 2 \sum_{j=1}^{n} \int_{p_{j-1}}^{p_j} \left(L_0(p_{j-1}) + \left(\frac{L_0(p_j) - L_0(p_{j-1})}{p_j - p_{j-1}} \right) (p - p_{j-1}) \right) dp \\ &= 1 - \sum_{j=1}^{n} (L_0(p_j) + L_0(p_{j-1}))(p_j - p_{j-1}) \end{aligned} \tag{7}$$

where $L_0(p_0) = 0$. Equivalence of (6) and (7) follows from

$$\begin{aligned} \frac{1}{T} \sum_{j=1}^{n} x_j (p_j + p_{j-1}) h_j &= \frac{1}{NT} \left(\sum_{j=1}^{n} x_j h_j \sum_{i=1}^{j} h_i + \sum_{j=2}^{n} x_j h_j \sum_{i=1}^{j-1} h_i \right) \\ &= \frac{1}{NT} \left(\sum_{i=1}^{n} h_i \sum_{j=i}^{n} x_j h_j + \sum_{i=1}^{n-1} h_i \sum_{j=i+1}^{n} x_j h_j \right) \end{aligned}$$

$$= \frac{1}{NT}\left(\sum_{i=1}^{n} h_i(T - TL_0(p_{i-1})) + \sum_{i=1}^{n} h_i(T - TL_0(p_i))\right)$$

$$= 2 - \sum_{j=1}^{n}(L_0(p_j) + L_0(p_{j-1}))(p_j - p_{j-1}).$$

Now suppose that the government dollar revenue requirement is C, so that $c = C/T$, and that we apply a linear income tax with marginal tax rate $r > c$. Then we have from (3) that

$$F(p_j) = rL_0(p_j) - (r - c)p_j$$

is the cumulative proportional tax through stratum j. Hence, for each stratum j, we have the following dollar amounts:

- Total Income Tax is

$$T(F(p_j) - F(p_{j-1})) = T\left(\frac{rx_jh_j}{T} - \frac{(r-c)h_j}{N}\right)$$
$$= (rx_j - (r - c)\bar{x})h_j.$$

- Average Income Tax per Household is

$$t_j = rx_j - (r - c)\bar{x}.$$

- Average Post-Tax Income per Household is

$$y_j = x_j - t_j = (1 - r)x_j + (r - c)\bar{x}.$$

- More generally, for any household with pre-tax income x , the linear dollar tax is $rx - (r - c)\bar{x}$, with percentage tax rate $r - (r - c)\bar{x}/x$. This is progressively increasing in x and approaches r for very large x.

At the other end of the scale, the tax is negative if

$$x < \left(1 - \frac{c}{r}\right)\bar{x},$$

the threshold for positive taxation. Moreover, we can use either (6) or (7) to verify (4) for the interpolated Ginis.

Using (6), we have

$$T(1 + G_0) = \sum_{j=1}^{n} x_j(p_j + p_{j-1})h_j$$

and

$$T(1 - c)(1 + G_1) = \sum_{j=1}^{n} y_j(p_j + p_{j+1})h_j$$

$$= \sum_{j=1}^{n}\left((1 - r)x_j + (r - c)\bar{x}\right)\left(p_j + p_{j-1}\right)h_j$$

$$= (1-r)T(1+G_0) + (r-c)\bar{x}\sum_{j=1}^{n}\left(p_j + p_{j-1}\right)h_j$$

$$= (1-r)T(1+G_0) + (r-c)T\sum_{j=1}^{n}\left(p_j^2 - p_{j-1}^2\right)$$

$$= (1-r)T(1+G_0) + (r-c)T$$

from which we obtain

$$G_1 = \left(\frac{1-r}{1-c}\right)G_0.$$

From the 2017 U.S. Census table [**14**], there are $N = 127.669$ million households and total annual household income of $T = \$11.189$ trillion, whereas average household income is $\bar{x} = \$87,643$. There are 42 strata with average household incomes ranging from \$1,128 to \$415,207. The pre-tax interpolated Gini index is $G_0 = 0.486$, close to the official Census figure of 0.482.

From the 2017 U.S. IRS table [**15**], we have total individual income tax revenue of $C = \$1.6053$ trillion. Combined with the Census household income data to simulate a general linear tax, we get a government tax ratio of $c = 0.14347 = 14.347\%$. With a marginal tax rate of $r = 0.50 = 50\%$, we have a common household subsidy of $(r-c)\bar{x} = \$31,248$ and a zero-tax threshold of $(1-c/r)\bar{x} = \$62,495$, below which income tax is negative. The average tax rate in the top stratum is 42% and the interpolated post-tax Gini index $G_1 = 0.284$, for an inequality reduction of 42%. So we see that significant inequality reduction is possible with a very simple linear tax structure. In Figure 4, we display pre-tax and simulated post-tax Lorenz curves based on the Census household income data, again showing significant inequality reduction in terms of twice the bounded area under the equity reference line.

The IRS table is intriguing beyond the total tax figure. There are 152.903 million individual tax returns and \$11.010 trillion total adjusted gross income (AGI), close to the Census total. There are 18 strata with average AGIs ranging from \$2,587 to \$31,259,606, a much more expansive range than the Census data. There is also a "No adjusted gross income" stratum which actually shows negative average AGI, although a relatively small number of returns pay a substantial amount of tax on average. We assume that this category is a mixture of reported capital losses on the one hand, and returns subject to the Alternative Minimum Tax on the other. For our purposes, this segment is simply ignored.

Beyond this wrinkle, the IRS table is subject to the same calculations. The interpolated pre-tax Gini index is 0.596, much higher than Census, reflecting a much more convex Lorenz curve. Moreover, the current income tax only reduces that to 0.561, a 6% reduction. In addition, the average tax rate in the top stratum is only 26%, lower than the previous few strata. Clearly, the current tax structure is not exacting on the high end and is not even progressive. We have not simulated a revised tax structure on the IRS table. In Figure 5, we display the pre-tax and post-tax Lorenz curves based on the 2017 individual tax return data. There is obviously very little inequality reduction in terms of twice the bounded area under the equity reference line.

Although the Census and IRS total pre-tax income figures are similar, the stratified tables and interpolated Lorenz curves are rather different. This difference stems from how the numbers are compiled and presented by the two bureaus. As indicated previously, the Census survey figures are widely disseminated and often quoted. Moreover, our interpolated pre-tax Gini is close to their official number. The IRS table, however,

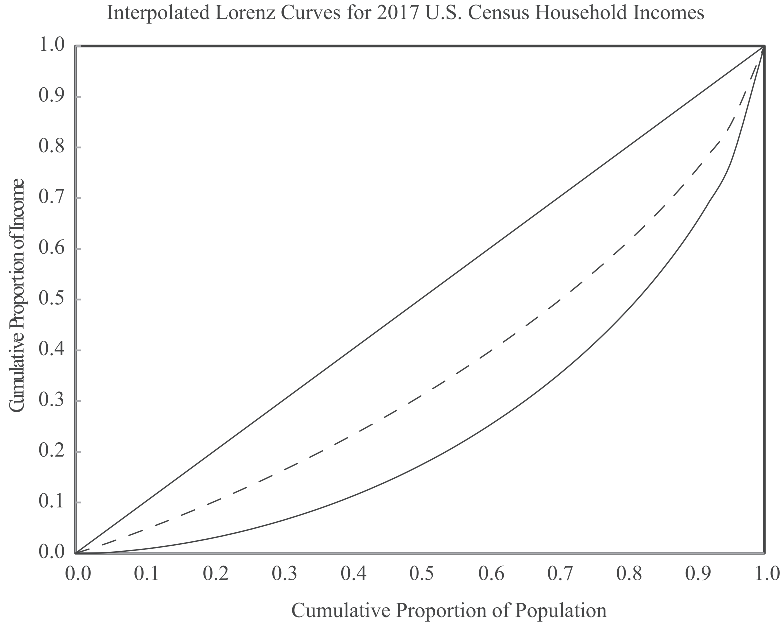

Figure 4 The solid line is the Pre-tax Lorenz curve. The dashed line is the simulated post-tax Lorenz curve.

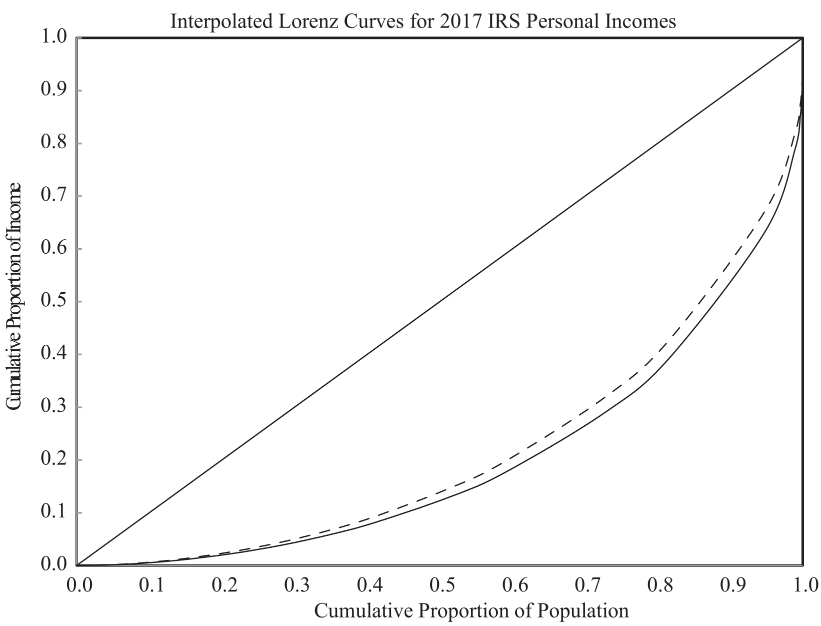

Figure 5 The solid line is the Pre-tax Lorenz curve. The dashed line is the Post-tax Lorenz curve.

is based on actual tax returns and explicitly covers a much broader range of incomes. We are not in a position to reconcile these two data sources, but we have found each to be individually useful in our research and they are corroborative in a general sense. We would like to thank both of these agencies for providing their data for analysis.

REFERENCES

[1] Aumann, R. J., Kurz, M. (1977). Power and taxes. *Econometrica*. 45(5): 1137-1161. doi.org/10.1007/BF02756484

[2] Diamond, P., Saez, E. (2011). The case for a progressive tax: from basic research to policy recommendations. *J. Econ. Perspect*. 25(4): 165-190. doi: 10.1257/jep.25.4.165

[3] Farris, F. A. (2010). The Gini index and measures of inequality. *Amer. Math. Monthly* . 117(10): 851–864. doi:10.4169/000298910X523344

[4] Fontenot, K., Semega, J., Kollar, M. (2018). *Income and Poverty in the United States: 2017*. U.S. Census Bureau.
www.census.gov/content/dam/Census/library/publications/2018/demo/p60-263.pdf.

[5] Friedman, M. (1962). *Capitalism and Freedom*. Chicago, IL: University of Chicago Press.

[6] Jantzen, R. T., Volpert, K. (2012). On the mathematics of income inequality: splitting the Gini index in two, *Amer. Math. Monthly*. 119(10): 824-837. doi.org/10.4169/amer.math.monthly.119.10.824

[7] Mankiw, N. G., Weinzierl, M., Yagan, D. (2009). Optimal taxation in theory and practice. *J. Econ. Perspect*. 23(4): 147-174. doi: 10.1257/jep.23.4.147

[8] Mirrlees, J. A. (1971). An exploration in the theory of optimum income taxation. *Rev. Econ. Stud*. 38(2): 175-208. doi.org/10.2307/2296779

[9] Piketty, T. (2014). *Capital in the Twenty-First Century*. Cambridge: Harvard University Press.

[10] Reich, R. B. (2015). *Saving Capitalism*. New York: Alfred A. Knopf.

[11] Saez, E., Zuchman, G. (2019). *The Triumph of Injustice: How the Rich Dodge Taxes and How to Make Them Pay*. New York: W. W. Norton & Company.

[12] Sheshinski, E. (1972). The optimal linear income-tax. *Review of Economic Studies*. 39(3): 297-302. doi.org/10.2307/2296360

[13] Stiglitz, J. E. (2012). *The Price of Inequality*. New York: Norton.

[14] United States Census Bureau. (2017). Table HINC-06, Income Distribution to $250,000 or More for Households: 2017. www.census.gov/data/tables/2018/demo/cps/hinc-06.html. Download 2017 Excel file is hinc06.xls.

[15] U. S. Internal Revenue Service. (2017). Table 1.1: All Returns: Selected Income and Tax Items by Size and Accumulated Size of Adjusted Gross Income, Tax Year 2017. www.irs.gov/statistics/soi-tax-stats-individual-statistical-tables-by-size-of-adjusted-gross-income. Download Excel file is 17in11si.xls.

Summary. Income inequality measurement via the Gini index is reviewed, both pre-tax and post-tax for a general tax curve. Focus then shifts to the linear income tax with a single marginal tax rate and a common subsidy for all taxpayers. It is demonstrated that significant inequality reduction is possible with this simple tax structure. Finally, we benchmark against real incomes data from the U.S. Census Bureau and the U.S. Internal Revenue Service.

ROBERT AGNEW (MR Author ID: 401319) received his PhD in applied mathematics from Northwestern University in 1968 under Erhan Cinlar. While on active duty in the United States Air Force, he taught operations research at the Air Force Institute of Technology. Subsequently, he pursued a business analytics career spanning retail, manufacturing, and financial services.