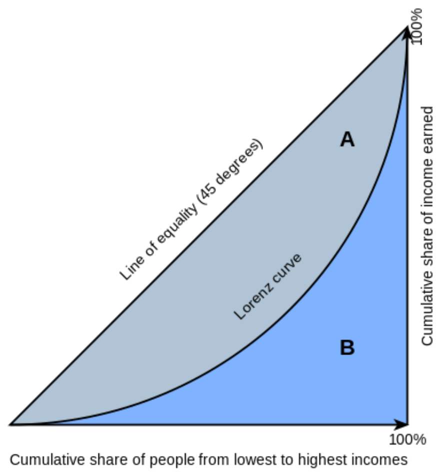
Income, Lorenz Curve, Gini Coefficients, and Taxes

Kason Ancelin

1 The Gini Index



1.1 Introduction to the Lorenz Curve

The Gini Index is a measure of income and/or wealth inequality among a population. It will be defined in a second, via a curve we plot called the Lorenz curve. It is of the utmost that you understand what the Lorenz curve represents as the rest of the packet is based on it.

As seen from the image, on the x-axis, we have the cumulative share of people from lowest to highest income. On the y-axis, we are plotting the cumulative share of income that these people have. You can read this as "people at or below the nth percentile of income all together receive y% of the total income". This is the point $(\frac{n}{100}, \frac{y}{100})$ on the Lorenz curve (the image above shows the x and y-axes with percentages, but we will have these axes span 0 to 1 instead of 0% to 100%). If this doesn't make sense right now, don't worry! The example below will illustrate the point and then come back and read this again to see if it makes sense.

This curve is best imagined with a discrete number of people. Imagine you have 10 people, and you put them in a line in the order of their incomes. The y-axis of the Lorenz curve is plotting the cummulative proportion of income as you move across the line from poorest to richest. Let's see a specific example.

Let's say we have a group of 10 people, with incomes as follows:

Person 1: 5

Person 2: 20

Person 3: 25

Person 4: 28

Person 5: 34

Person 6: 56

Person 7: 80

Person 8: 82

Person 9: 100

Person 10: 200

The total income of all these people is 630.

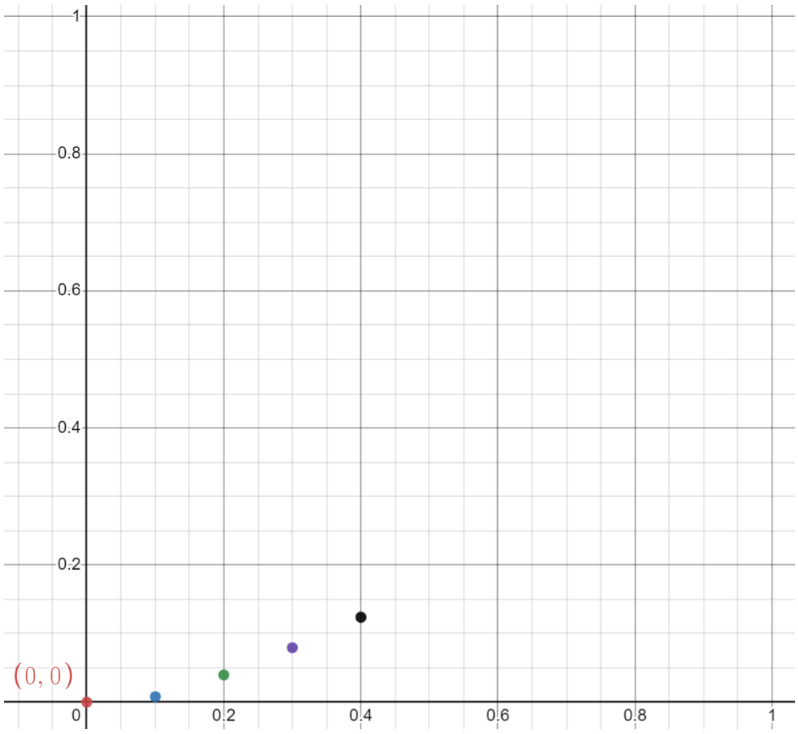
At the 10th-percentile of income (i.e just person 1), the cumulative sum of income of everyone at or below this percentile is 5. So, everyone at or below the 10th-percentile of income all together have $\frac{5}{630} = \frac{1}{126}$ of the income in the country. On the Lorenz curve, this corresponds to point $(0.1, \frac{1}{126})$.

At the 20th-percentile, the cumulative sum of income of everyone at or below this percentile is just the sum person 1 and person 2's incomes, which is 25. So, everyone at or below the 20th-percentile of income all together have $\frac{25}{630} = \frac{5}{126}$ of the income in the country. On the Lorenz curve, this corresponds to point $(0.2, \frac{5}{126})$.

At the 30th-percentile, the cumulative sum of income will be the sum person 1, 2, and 3's incomes, which is 50. Thus, everyone at or below the 30th-percentile of income all together have $\frac{50}{630} = \frac{5}{63}$ of the income in the country. On the Lorenz curve, this corresponds to point $(0.3, \frac{5}{63})$.

At the 40th-percentile, the cumulative sum of income will be the sum person 1, 2, 3, and 4's incomes, which is 78. Thus, everyone at or below the 40th-percentile of income all together have $\frac{78}{630}$ of the income in the country. On the Lorenz curve, this corresponds to point $(0.4, \frac{78}{630})$.

We have plotted the points on a graph below, beginning the process of constructing the Lorenz curve from the computations we completed above. Now, the rest follows quite easily and is for you to complete. Compute the rest of the cumulative shares of income at each percentile level (i.e at each interval of 10 percentile level) just as above and plot them on the Lorenz curve. Then, just connect the dots with lines so that it is a "curve". What do you notice about this curve? We will soon talk about some facts about the Lorenz curve and proof them!



Problem 1.1.

Explain why the Lorenz curve is always increasing.

Problem 1.2.

Explain why the line segment ($y = x$) from $(0, 0)$ to $(1, 1)$ should be called the "equal distribution line".

Problem 1.3.

Explain why every Lorenz curve is bounded below the equal distribution line. Think of what it would mean for a point on the Lorenz curve to be above this line and derive a contradiction.

Problem 1.4.

Assuming no one in a population earns a negative income, explain why the Lorenz curve must start at $(0, 0)$ and end at $(1, 1)$.

Problem 1.5.

Assuming no one in a population earns a negative income, show the only possible Lorenz curve having area $1/2$ below the graph (i.e in region $A+B$) is the equal distribution line. (Hint: Use that the Lorenz curve is bounded below the equal distribution line)

Problem 1.6.

Which of the following functions defined only on the interval $[0,1]$ can be Lorenz curves? Draw them to see if needed.

$$f(x) = x$$

$$f(x) = x^2$$

$$f(x) = x^3$$

$$f(x) = x(1-x)$$

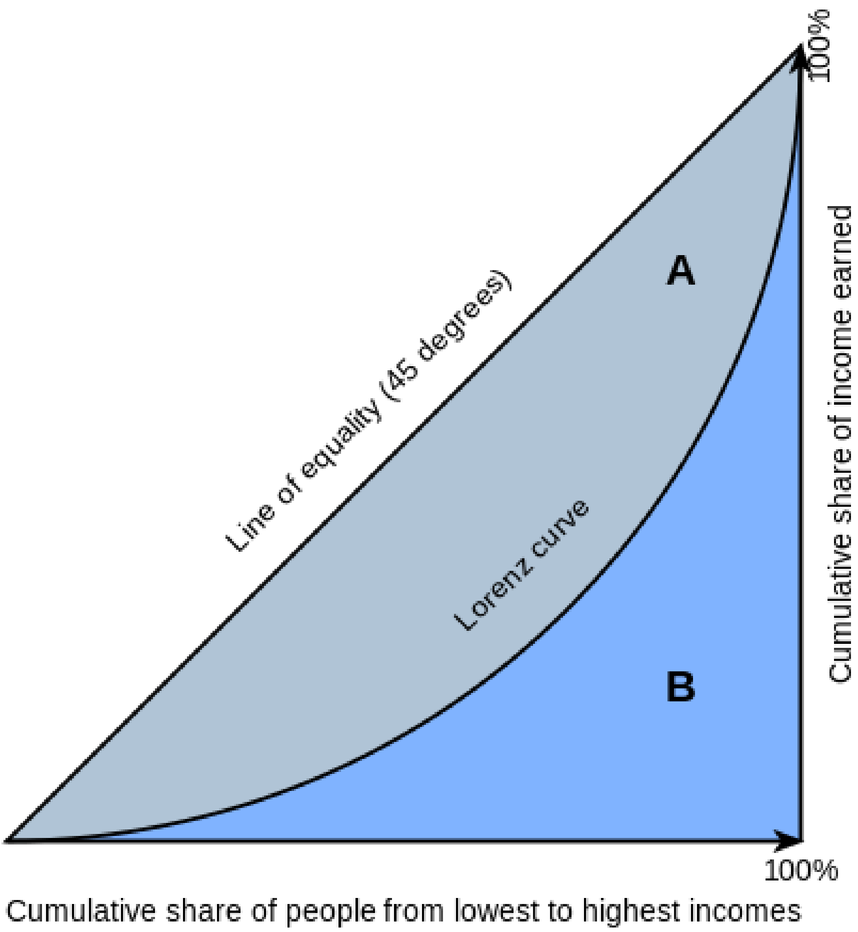
$$f(x) = x + 1/2$$

$$f(x) = \sqrt{x}$$

$$f(x) = e^x - 1$$

Note that Lorenz curves can be piecewise linear (i.e lines with positive slope stitched together) or smooth (like x^2). In discrete cases, it will be piecewise linear while in continuous cases (where we assume a ton of people in a population and thus the curve is very smooth) we may use a general smooth Lorenz curve.

1.2 Intro to The Gini Coefficient



With the Lorenz curve in hand, the Gini coefficient is really easy to define! The Gini coefficient can be defined as the ratio of the area that lies between the line of equality and the Lorenz curve (marked A in the diagram) over the total area under the line of equality (marked A and B in the diagram); i.e., $G = \frac{\text{area of } A}{\text{area of } A \text{ and } B}$. For the rest of this packet, we will refer to the area under the Lorenz curve as B, as indicated in the figure above.

Problem 1.7.

Prove that if there are no negative incomes in a population, then the Gini coefficient, G, is equal to $2a$, with a denoting the area of region A. (Hint: Think about what the area of region $A + B$ is.)

Problem 1.8.

Prove that if there are no negative incomes in a population, then the Gini coefficient, G, is equal to $1 - 2b$, with b denoting the area of region B.

For the rest of this packet, we will assume there are no negative incomes (not because that's mean, but actually because the math loses some extraordinary properties, and it is quite a reasonable assumption).

Problem 1.9.

What is the range of possible values the Gini coefficient can take?

Problem 1.10.

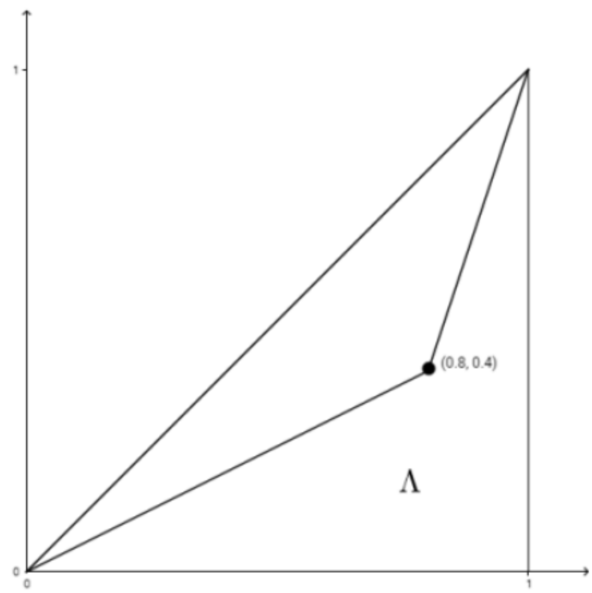
Finish the statement: A larger Gini coefficient close to 1 represents a (large/small) amount of income inequality whereas a smaller Gini coefficient close to 0 represents a (large/small) amount of income inequality.

Now we understand that the Gini coefficient is a descriptive statistic for income and/or wealth inequality among a population.

We now turn to problems to explicitly compute some Gini coefficients (note that the Gini index is the same thing as the Gini coefficient). Recall that we call B the area marked Λ on the pictures below. (Source for problems 11 through 15: Dillon Zhi, The Gini Index Handout)

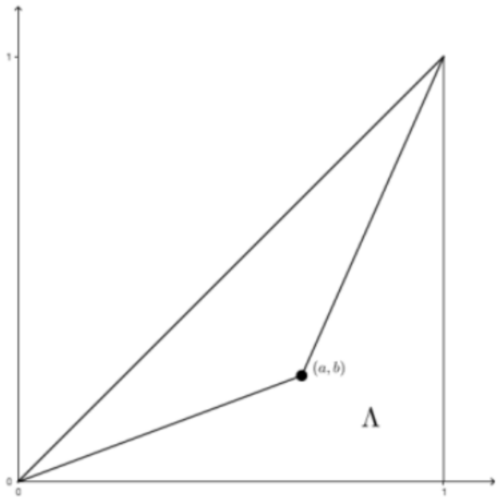
Problem 1.11.

Given the Lorenz curve consisting of two line segments in the figure, calculate the area of B and use that information to compute the Gini index.



Problem 1.12.

Calculate the area of B in the figure and then write a general formula for this "one-point estimate" of the Gini index in terms of a and b.

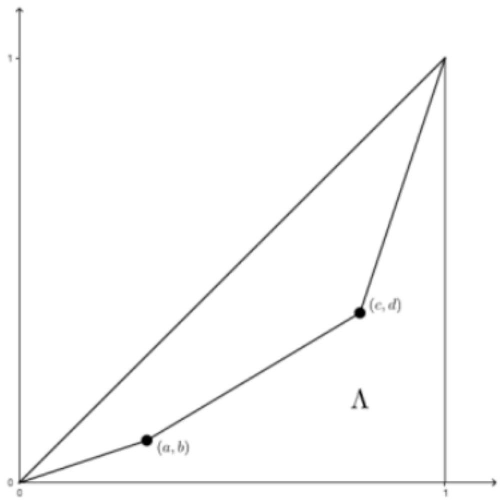


Problem 1.13.

If we have our calculation of the Gini index on a one-point estimation, will that estimate be larger or smaller than the actual value of the Gini index? Why?

Problem 1.14.

Calculate the area of B in the figure and then write a general formula for this "two-point estimate" of the Gini index in terms of a, b, c, and d.



Problem 1.15.

Use the answer to the previous problem to make a two-point estimate of the Gini index if $(a, b) = (.8, .4)$ and $(c, d) = (.99, .8)$.

2 Taxing Time

It is time for taxes. This section serves to show a variable which can alter a Lorenz curve. We define a regressive tax system as one where those make more pay a smaller percentage of their earnings and a progressive tax system as the one where those who make more pay a larger percentage of their earnings. More detail could be added to these definitions, but it is not needed for the following qualitative problem.

Problem 2.1.

For a population with income tax, we redefine everyone's incomes to be their incomes after tax, and then construct the Lorenz curve with these post-tax incomes. Draw an initial Lorenz curve for the economy before taxes. Now depict how the Lorenz curves changes upon introducing a taxing system. First, draw the new Lorenz curve for an economy when a regressive taxing system is implemented. Then draw the new Lorenz curve when a progressive taxing system is implemented. How do these system's alter the

Gini coefficient?

We could make you do calculations using these taxing systems, real Lorenz curves, and calculating the change in Gini coefficient, but we will leave this for another time.

3 Game Time

This section will be completed together after the break. Everyone has the chance to earn Sanjit bucks in these games, which everyone from last school years class knows are quite valuable. If you got here early, just continue on past this game section and we'll revisit it together after the break.

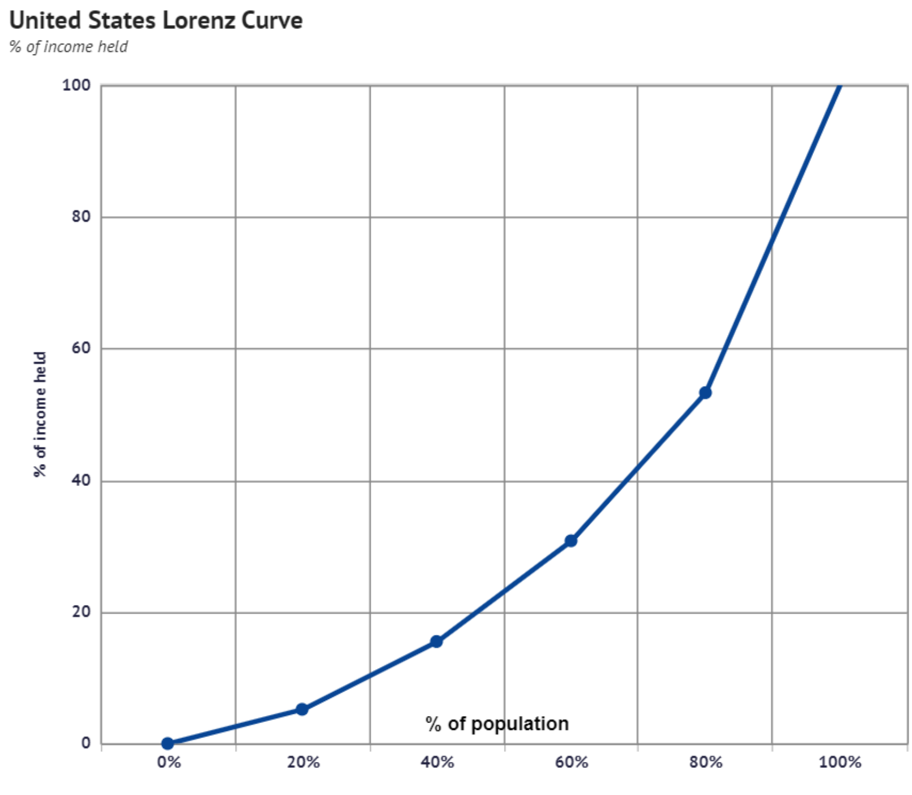
3.1 Game 1: The Warmup

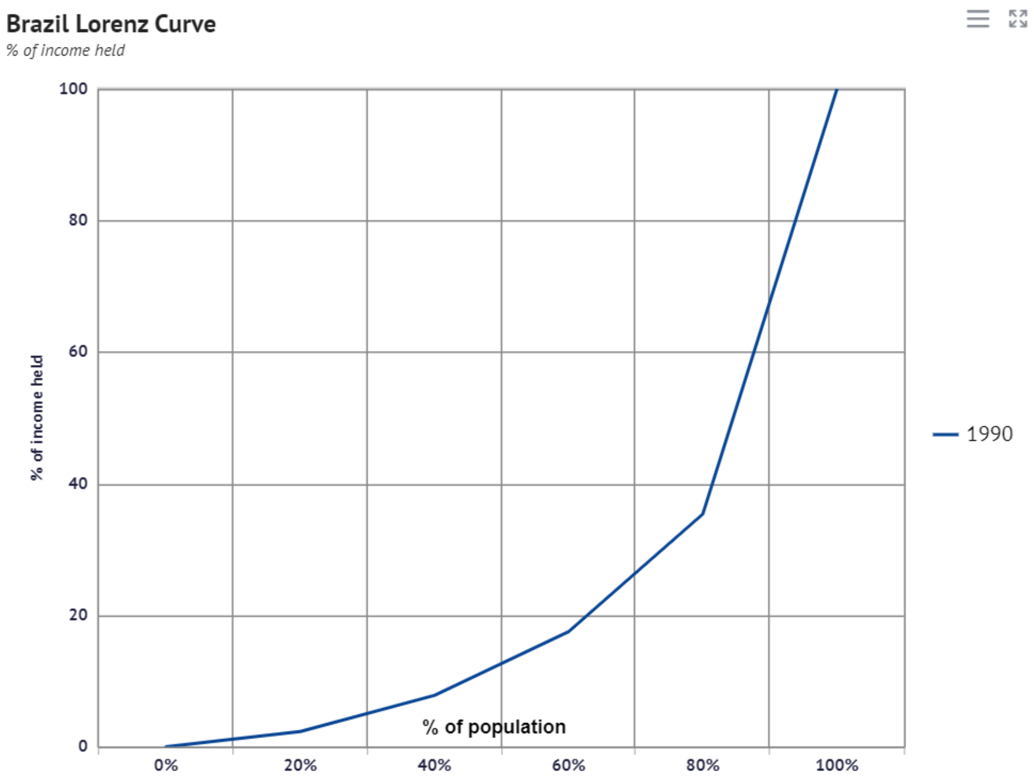
In Ohio, let's say there are 100 people. You will pick an integer n between 1 and 100 inclusive. You get paid $100 - n$ plus the cumulative income of all people at or below the nth percentile of the population. You also know the total income among the population is 100. Pick your n. A follow up: does your answer change if Ohio has many more than 100 people?

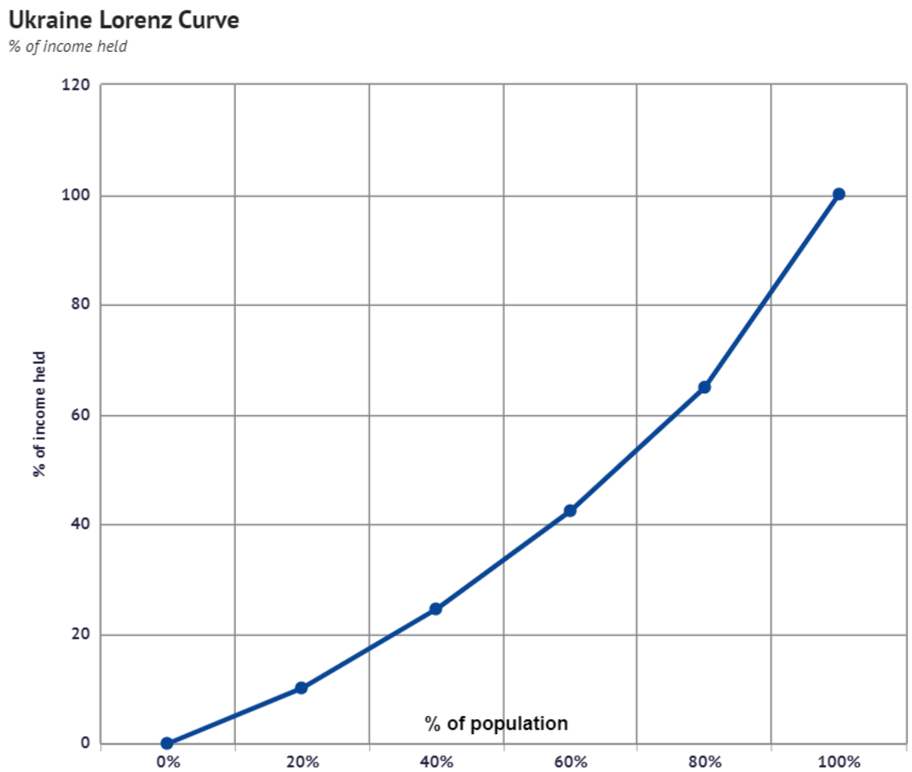
3.2 Game 2

We are again in Ohio. Your instructors have picked a specific Lorenz curve for Ohians. You will pick an integer n between 1 and 100 inclusive. You will pay $2^{n/10}$ to obtain the cumulative income of all people at or below the nth percentile of the population. You also know the total income among the population is this time 1000. Pick your n.

4 Real Lorenz Curves







Problem 4.1.

Rank the Gini coefficients of the three countries above in order. To find tons of Lorenz curves and Gini coefficients using real data from various countries, one can go to https://knoema.com.

This is where these graphs were pulled from.

5 Challenge Problems

Problem 5.1.

You are given that the Gini coefficient of Ohio is 0.95. In this economy, the top 1% make more than the bottom n%. Find the largest n such that we can guarantee this statement is true. (Hint: Think about extreme cases to minimize or maximize areas)

Problem 5.2.

Suppose there are 100 people in a population and you are given a Lorenz curve for wealth. How could you determine the nth person's wealth? (Hint: Think on the margin like an economist, maybe revisit the explicit example done earlier in this packet with 100 people.

Geometrically, it may involve slopes)

Problem 5.3.

Now let's refine the result above. Suppose there are $350,000,000$ people in a population and you are given a Lorenz curve for wealth. How could you roughly determine someone in the pth percentile's wealth? (Hint: It is virtually the same argument as above, except now with many more people, the graph has become much smoother and less rigid.)

What do the above problem's tell us? They tell us that the wealth of an individual at the pth percentile is roughly the "slope" of the Lorenz curve times the total wealth in the population. In fact, one can rigorize this using calculus when we assume a Lorenz curve with an infinite population, but this is behind the scope of our work today. If you are interested, feel free to ask an instructor about this!