Income Inequality: Employing the Gini Coefficient

Katelynn Comer

University of North Georgia, GISC 3200 Fall 2020, Huidae Cho

Abstract

This paper reviews the importance of understanding income inequality and address gaps in our spatial understanding of it. It presences a systematic way of quantifying this inequality and makes to useful for spatial analysis. It also discusses the importance of examining every aspect of income inequality analysis.

1. Introduction

i. Measures of Income Inequality:

There is a variety of techniques used by researchers to measure income inequality. The use of indices appears to be favored by most, but let us discuss a few methods, their advantages and their weaknesses. A common measure is the Theil Index, or mean log deviation. In this index the average distance of between median and average income or the mean logarithmic deviation of income (MLD) is calculated [1]. An advantage of this type of measurement is that is completely decomposable, i.e. can be broken down by characteristics such as population groups like age or race, rural or urban areas, income sources, etc. This allows the index to have the ability to quantify what portion of income inequality is a result of differences within and across groups. If the index is equal to zero perfect equality is represented by the population, as the index increases the distribution is more unequal. A short coming of the Thiel Index is that the coefficient is not capped. This means that the measurement is not relative and therefore is not always comparable across population of different sizes [2]. Lacking this quality does not make this a good candidate for our analysis.

Atkinson's index is popular analysis for welfare-based measures of inequality. In his *text The Economics of Inequality* Atkinson stated that inequality "cannot, in general, be measured without introducing social judgements. [Other measures] are not purely 'statistical' and they embody implicit judgements about the weight to be attached to inequality at different points on the income scale" (pg. 47) He worried that other measures of inequality did not allow for varying sensitivities across the income distribution. To resolve this concern his calculation includes a unique parameter: the degree of society aversion to inequality [3]. This parameter is theoretical and decided by the researcher. This number can range anywhere between zero and infinity where a larger value implies the willingness of a society to accept lower wages to promote equal income distribution or, more accurately, the researcher's interpretation of said willingness. Like the Theil Index, the Atkinson's Index allows for the break down into smaller groups to allow for comparison within and across groups. However, it is criticized for not having an obvious or intuitive interpretation and therefore is rarely cited outside of academic journals [5].

Another method is to consider the proportion of total income earned. Using this approach, we examine the poorest percentile of the population and is considered one of most intuitive measures for examining income inequality. Most commonly used ratios for this analysis are 90^{th}% to 50^{th}%, representing the gap between the rich and the middle classes; 90^{th}% to 10^{th}%, representing the gap between the rich and bottom classes; and 50^{th}% or 10^{th}%, representing the gap between the middle and the bottom classes [4]. A notable limitation of this logic is the limited insight to actual income distribution of a population. The percentage of income earned by the bottom 50^{th}% does not inform the researcher how evenly the income is earned across the population [6].

The most used measure for income inequality is the Gini coefficient, which is directly connected to the Lorenz curve. The Lorenz curve is one the simplest methods for representing inequality. It is a curve on a graph with x-axis: the income recipients ranked poor to rich and a y-axis: percentage of total income. In this technique perfect equality is represented by a 45˚ line, this occurs when all income is shared equally among members of the population [2]. As inequality increases the Lorenz curve moves further away for this line of equality.

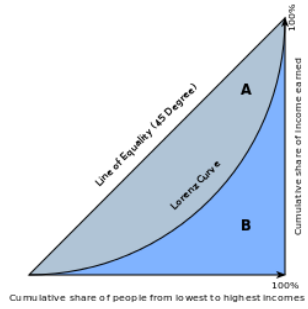

FIG 1: Lorenz Curve

On its own the Lorenz curve cannot conclusively determine the level on inequality of the sample population. The Gini Index is necessary to accomplish this. The Gini Index measures "the extent to which the distribution within an economy deviates from a perfectly equal distribution" [3]. It is a complex calculation developed in 1912 by the Italian statistician Corrado Gini. It can be used or without the associated Lorenz curve.

$$G = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n}|x_i - x_j|}{2\sum_{i=1}^{n}\sum_{j=1}^{n}x_j} = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n}|x_i - x_j|}{2n\sum_{j=1}^{n}x_j} = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n}|x_i - x_j|}{2n^2\bar{x}}$$

FIG 2: Gini Coefficient Equation

However, if we take a logical approach to how Gini is calculated we understand that it is a ratio between the two curves (labeled 'A' in Fig. 1) and the entire area under the curve of perfect equality ('A' + 'B' in Fig. 1) [3]. Therefore we can think about the Gini index as:

$$\text{Gini} = A / (A + B)$$

FIG 3: Gini-Logical Approach

The higher the Gini coefficient is, the more unequal distribution of income is occurring. One major disadvantage in using this calculation is that is it not easily decomposable like others, as it examines income. However, unique to the Gini index is that it is independent of the population or income size. This means it can be used to directly compare across different populations' income distribution [3]. This aspect of the Gini Index is what makes it so valuable and prolific in studies regarding income inequality and furthermore, why it has been chosen for this review. Note that the Gini index solely evaluates income distribution and does not account for income level, i.e. rich nations can have the same Gini index and poor nations. Note also that the Gini index for a region will change depending on income definitions. For this reason, it is important to understand what constitutes income.

ii. Income:

What constitutes income? What seems like a basic question is unknowingly quite complex and vital to evaluations of income inequality. The traditional Schanz-Haig-Simons income definition is "is the value of consumption plus change in net wealth" [5]. However, a more modern and common definition is the money received over a year usually as payment for work. By this definition money can include any of the following: wages and bonuses, investment returns and business payments, rental property income, elderly assistant such as pensions and social security, etc. [5].

Several economists criticize this definition and believe it is not inclusive enough. They suggest that employer-provided benefits should be classified as income of those who receive them. Proponents of this view argue that these benefits, health insurance and retirement accounts for example, "directly benefit workers and help shape current and future standards of living" [5].

While this is certainly true, not everyone agrees with this logic, explaining that these contributions are out of the employees control therefore, should not be included. They further argue that, with using this logic, government services and payments should be considered. This includes medical services performed at public clinics and veterans' hospitals, food stamps, housing vouchers, etc. [5]. Others complain that appreciation of assets further convolutes income measures. In tax returns, for example, capital payments are calculated but change in home values, stocks, and interests are ignored. While others account for all asset gains to track yearly wealth [7]. So, on and so forth.

Instead of falling down this slippery slope logic, we can examine an organization who has tackled this problem of defining income and tacking it in populations across the United States, the U.S. Census Bureau. The Current Population Survey (CPS) is the largest, oldest, and well-known survey conducted in the United States [8]. Through its collection, it provides information about individuals such as occupation, earning, education, etc. Also included are comprehensive and detailed questions concerning income and income sources. For every person over 15 years old the CPS asks questions from each of the following categories: earnings; unemployment compensation; workers' compensation; social security; supplemental security income; public assistance; veterans' payments; survivor benefits; disability benefits; pension or retirement income; interest; dividends; rents, royalties, estates and trusts; educational assistance; alimony; child support; financial assistance from outside the household; financial; other [1]. It is important to note these measurements are taken before taxes, social security, union dues, Medicare payments etc. Because of this decision by the CPS and Census Bureau, non cash benefits are also not reflected in the income calculation, such as subsidized housing, SNAP/ food stamps, goods produced and consumed on the farm, etc. [1]. Additionally, the Census Bureau informs us that "readers should be aware that for many different reasons there is a tendency in household surveys for respondents to underreport their income" [1]. While, some level of error is impossible to avoid, especially when considering the human element, it is still important to note. Despite this fact, the CPS's goal is to provide detailed and timely income and poverty estimates and is used as an official source in the national poverty estimates [1].

2. METHODS

i. Field of Study

Questions from the Current Population Survey are broken into three geographic categories: households, families in households (called just 'families'), and people in households. Households are considered every person living in the same residential unit. Excluded from this group are those who live in group settings, areas like military bases, care facilities, jails, etc. [5]. Household income is the sum of every income in that unit, however, families in households excludes anyone not related to the homeowner who resides in the household. There is minimal difference between these two groups, with only 6% of people not being related to the householder [5]. Median household income is the most widely used measure of income when performing statistical analysis. This is because it considers all households in the geographic region and is seen as an accurate representation of typical income level in that area.

Furthermore, the Current Population Survey is broken down by geographic regions. Five-year estimates are available for all geographic units such as census tracts and block groups [1]. Census blocks are the smallest unit surveyed by the Census and began in the 19040s [10]. Its origin was for the operational needs of the Census Bureau. The creation of blocks allowed for data collection to be sequenced in an efficient way [10]. Their assemblage is far more complex in modern day, however. While it is intuitive to regard census blocks as regular, consistent, regular or square, the configuration is actually far from. Many factors influence the shape decision of blocks and vary greatly between urban and rural areas as well as regional across the US. Our area of focus will remain in Georgia. Rural areas are highly influenced by topography and which

survey system was used at the time of settlement [10]. In most urban areas you'll find a grid system of small blocks with distribution because of topography, water features, and special land use areas. Older suburbs reflect this pattern, while newer suburbs tend to have larger blocks to accommodate the curved street patterns [10]. Next in the geographic hierarchy is block groups, consisting of up to nine homogenous blocks. Block groups typically consist of between 600 and 3000 people [11]. Block groups are a good sample size for studying inequality because it is a small enough area, we can have a complete view of income distribution but large enough were they can still be processed efficiently.

ii. Data Collection

Data collected by the census and published by the National Historical Geographic Information System was used [12]. A python tool was designed using Esri's ArcGis Pro.

iii. Problem Solving

To address this gap in ArcGis Pro software a "Calculate Inequality" tool to employ the use of the Gini Coefficient has been created. The first step was to decide how to perform the algorithm. Taking inspiration from the Riemann sum style of approximation, a loop was designed. By using repeating triangles and rectangles to approximate the area under the Lorenz curve, the Gini coefficient can be calculated from a list of values (income per census block group in this case).

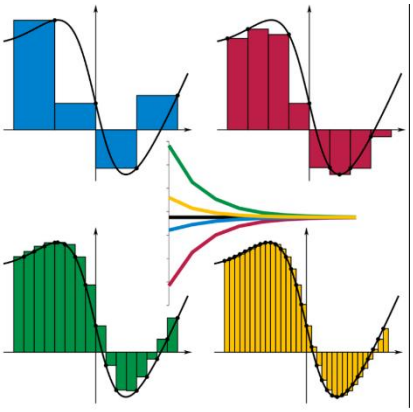

FIG 4: Riemann Sum Visualization

After this was created, preexisting tools were used to create seamless workflow. Specifically, the calculate field tool, which offers a code block area in which the loop designed above was utilized.

- #1 Join income information to GA block shapefile
- #2 Add New Field
- #3 Use Calculate field to determine Gini Coefficient
- #4 Find county wide statistics
- #5 change symbology to display Gini Coefficient

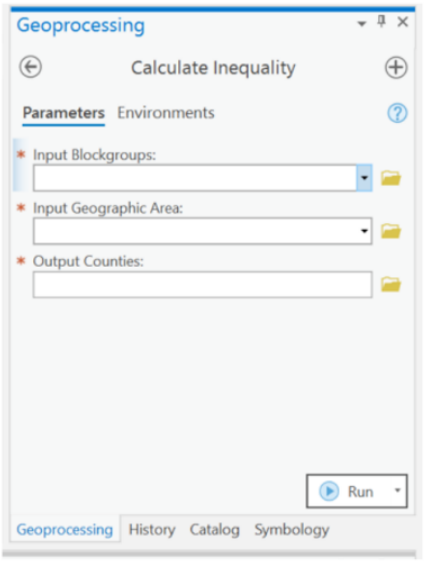

FIG 5: Workflow Breakdown

Note: In its current state the tool does not consistently automatically change the symbology and must be done manually on occasion.

3. RESULTS

This is an example of two analysis that were ran using the Calculate Inequality Tool:

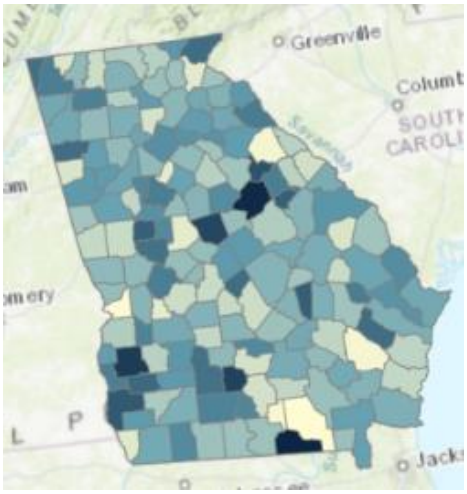

FIG 6: Georgia Income Inequality 2010

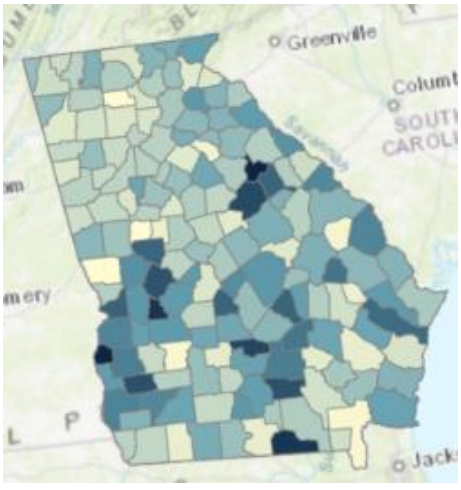

FIG 7: Georgia Income Inequality 2018

The darker the county appears the higher its Gini coefficient. When comparing occurs almost a decade of time we see a trend toward higher income equality. The tool also offers flexibility in its study area. Any geographic can be analyzed given, you have the block group data. We could use state boundaries in replace of county boundaries and if we wanted to compare across states. For the first time we have a way to calculate the Gini coefficient and analyze it in the spatial dimension in one step. By creating a standard unit of measure, we can detect patterns and use this tool as a launching pad for further research. Understanding income distributions can give indication of the economic well-being of areas. It can also give policy makers insight in the communities of areas and guide decision making.

References:

1. DeNavas-Walt, C., & Proctor, B. (2014). Income poverty and inequality. *U.S. Department of Commerce Economics and Statistics Administration*. doi:10.18356/722a00c5-en
2. Trapeznikova, I. (2019). Measuring Income Inequality. *IZA Institute of Labor Economics*. doi:10.15185/izawol.462
3. Afonso, H. (2015). Inequality Measurement Development Issues No. 2. *Department of Economic and Social Affairs*. Retrieved 2020, from https://www.un.org/en/development/desa/policy/wess/wess_dev_issues/dsp_policy_02.pdf
4. Atkinson A B. The economics of inequality. Oxford: Clarendon Press, 1975
5. Rose, S. (2018). Measuring Income Inequality in the US Methodological Issues. *Income and Benefits Policy Center*. Retrieved 2020, from https://www.urban.org/sites/default/files/publication/99456/measuring_income_inequality_in_the_us_methodological_issues_final.pdf
6. De Maio, F. (2007). Income inequality measures. *J Epidemiol Community Health*. doi:10.1136/jech.2006.052969
7. Larrimore, J., Burkhauser, R., G. A., & Armour, P. (2017). Recent Trends in U.S. Top Income Shares In Tax Record Data Using More Comprehensive Measures of Income Including Accrued Capital Gains. *National Bureau of Economic Research Working Paper Series*. Retrieved 2020, from https://www.nber.org/system/files/working_papers/w23007/w23007.pdf
8. Bureau, U. (2019, February 11). About this Survey. Retrieved December 05, 2020, from https://www.census.gov/programs-surveys/cps/about.html
9. (n.d.) All About Measures of Income in the Census. Retrieved December 05, 2020, from https://mcdc.missouri.edu/help/measures-of-income/
10. Census Blocks and Block Groups. (n.d.). Retrieved 2020, from https://www2.census.gov/geo/pdfs/reference/GARM/Ch11GARM.pdf
11. Lemery, C. (2020). United States Census Information: Understanding Census Geography. Retrieved 2020, from https://pitt.libguides.com/uscensus/understandinggeography