Discrimination Assignment - I

Aftab Ahmed, Bishmay Barik, Manya Pahwa

3rd November 2024

1 Introduction

This document provides a brief explanation of how the Gini Coefficient heatmaps were generated and how the ELMO measure for different religions and castes was calculated, with results included at the end. This document also includes two appendices. In Appendix 1, we present heatmaps for Gini Coefficients. Appendix 2 contains brief documentation explaining the contents of each folder and file in the `Discrimination_Assignment_ABM.zip` archive.

1.1 Calculation of MPCE and Heatmaps

In this assignment, our group selected Tamil Nadu as the focal region for analyzing income inequality by calculating the Gini coefficients of Monthly Per Capita Expenditure (MPCE) at the district level. The accompanying STATA do-file, provided with this document, details the data preparation, cleaning, and analytical procedures employed.

We began by processing the Household Consumption Expenditure Survey data, resolving inconsistencies, and applying sampling weights to compute the MPCE for each district in Tamil Nadu. A complete description of this procedure is documented in the STATA do-file.

The computed Gini coefficients for MPCE were visualized as a district-level heatmap for Tamil Nadu. The base map used for this visualization was sourced from this resource. The calculated Gini coefficients are displayed in the subsequent graphs.

District-level Gini coefficients were computed in STATA, and to ensure accurate alignment between districts in the Household Consumption Expenditure Survey data and the Tamil Nadu district shapefile, we employed a fuzzy matching technique in Python using the `fuzzywuzzy` library. The Python code used for this matching is included with this document.

The Appendix contains two visualizations. In Graph A, we present the Gini coefficients across districts as a continuous gradient, with districts shaded in varying intensities of orange to reflect levels of income inequality. Additionally, Graph B categorizes the Gini coefficients into discrete bins, based on the minimum and maximum Gini values observed across districts.

1.2 Calculation of ELMO measure for different religion and caste categories

Further, we calculated the ELMO measure as given by (Elbers, Lanjouw, Mistiaen, & Özler, 2008) to estimate inequality accruing to religion and caste structure separately, in the rural Tamilnadu. The formula for ELMO measure is:

$$\hat{R_B}(y) = \frac{I_B(y)}{\max\{I_B|y, G, \frac{n1}{n}...\frac{n_G}{n}\}} \tag{1}$$

where, I_B refers to 'between group inequality' and y stands for MPCE. The total number of groups is represented by G and $\frac{n_G}{n}$ is the relative size of group G.

In this assignment, we followed these steps to calculate the ELMO:

1. **Average MPCE for each subgroup:** We first calculated the average MPCE for each sub group. For example, while calculating the ELMO measure for caste, we calculated the average MPCE for SC, ST, OBC and General categories in the first step.

2. **Sorting by MPCE:** Then, we arranged all individuals in ascending order of their MPCE.

3. **Maximising the between group inequality (Denominator):** We then reassigned the individual group identities by allocating the group with the lowest average MPCE (with relative size x) to the (x) individuals with the lowest MPCE. Similarly, we assign the group with the second lowest average MPCE to the individuals with second lowest MPCE and so on. In other words, we keep the relative group size and the order of average MPCE of the respective groups the same, while rearranging. This is how the between group inequalities are maximised.

4. **Average MPCE for the rearranged groups:** Next, we calculated the average MPCE for the rearranged groups.

5. **Using Theil's L to calculate ELMO:** Finally, we use Theil's L as the measure of inequality and calculate the same for the original (caste/ religion groups) and rearranged groups and plug them in the formula given above to arrive at the ELMO measure. Theil's L for the original groups (I_B) makes the numerator, while Theil's L for the rearranged groups (max I_B) makes the denominator.

1.3 ELMO Measure Results

First, we have calculated the ELMO measure for rural districts in Tamil Nadu, across all religions. The ELMO measure for religions we have achieved is **.1012587**. Next, we have also calculated the ELMO measure for caste categories in rural districts of Tamil Nadu. The value which we have received is **.0529528** .

2 Conclusion

Through this assignment, we have solved for the district-wise Gini Coefficients for MPCE across rural districts in Tamil Nadu. The results which we have achieved for Gini Coefficients have been submitted in the zipped folder (please refer Appendix - 2 for more details). Further, we have also calculated the ELMO measure for rural districts in Tamil Nadu, for both religion and caste categories. The ELMO measure for religion is .1012587 and for caste categories, it is .0529528. The codes for calculating these ELMO measures can be found in the STATA codes provided, in `05_ELMO.do`

Appendix 1: Graphs

Graph A

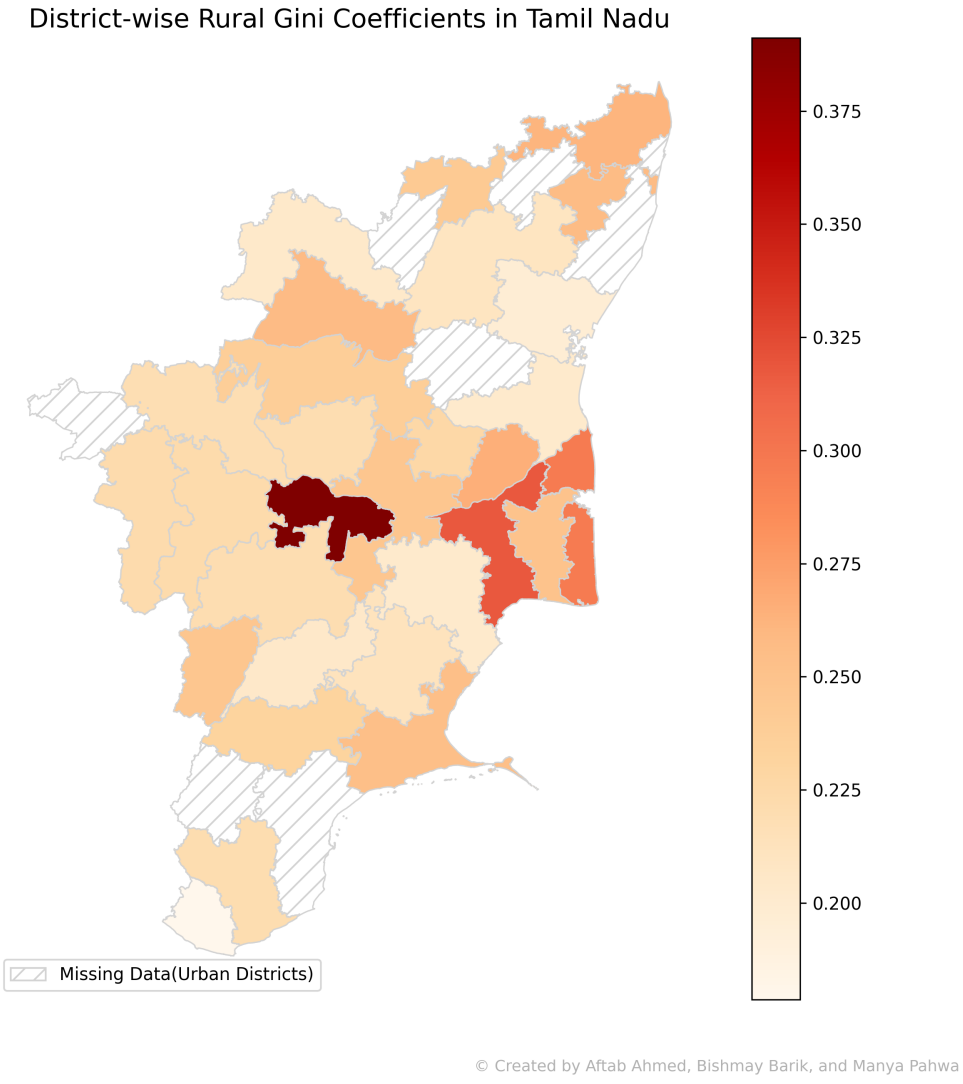

Figure 1: Gini Coefficient Map for Rural Tamil Nadu (Continuous)

Graph B

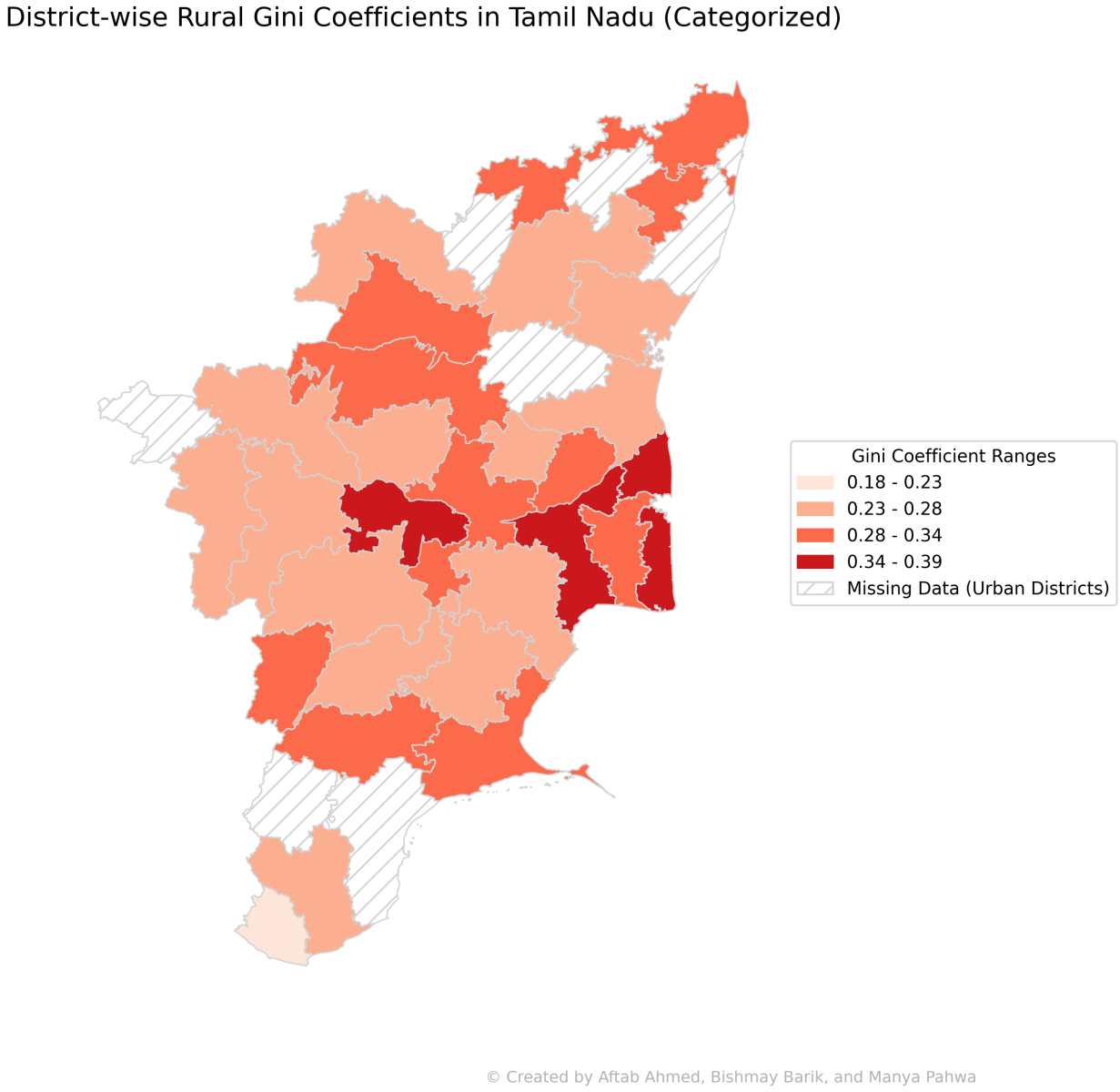

Figure 2: Gini Coefficient Map for Rural Tamil Nadu (Categorized)

Graph C - The STATA Graph

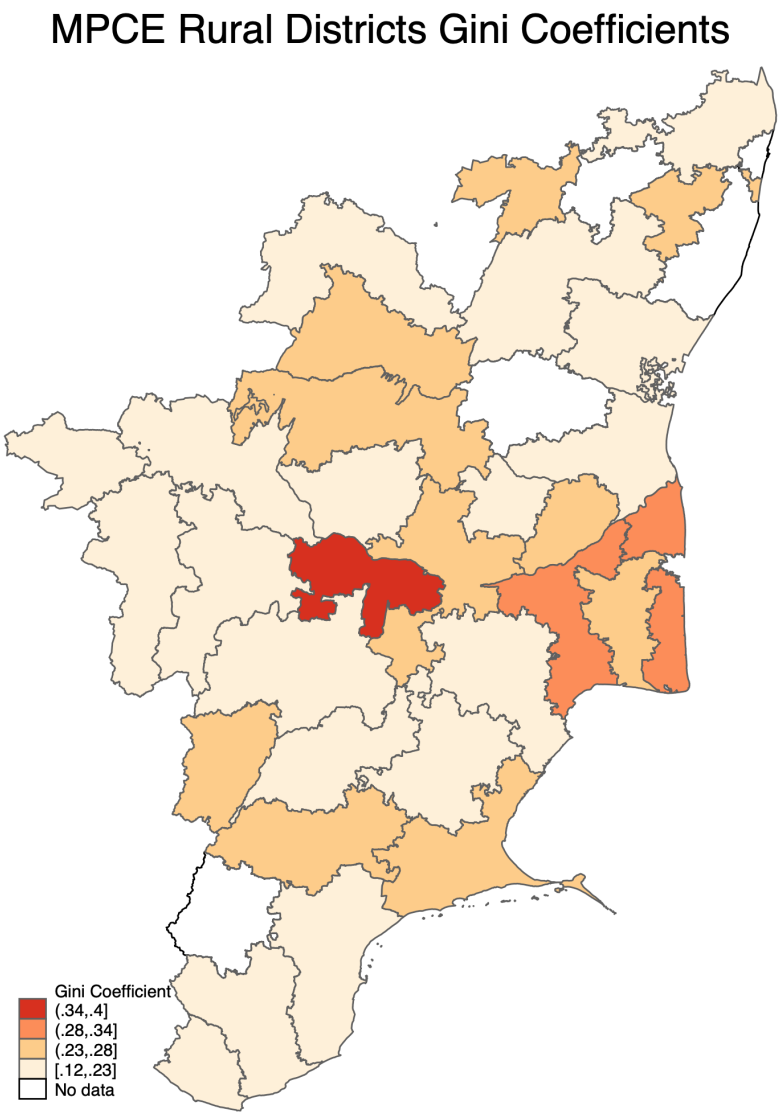

Figure 3: Gini Coefficient Map for Rural Tamil Nadu (Categorized)

Appendix - 2 (Documentation)

The name of the zipped file which we have shared with you is `Discrimination_Assignment_ABM.zip`. Once the file is uncompressed, you'll have the unzipped folder with you. `Discrimination_Assignment_ABM` would contain the following folders:

- `01. Raw`: This will consist another folder named `TXT`, which contains the raw TXT files of HCES data. The files have been downloaded from here.
- `02. Output`: This output folder contains five more folders within it, namely:
  - `Shapefile`: This contains all the shapefiles which we have downloaded from various sources. We have mainly used the shapefile which is present in the `TN_Shapefile` folder. This can be downloaded from here.
  - `dta`: This folder contains the dta cleaned versions of the HCES data.
  - `csv`: This folder contains the csv cleaned version of the HCES data
  - `mpce`: This folder contains the mpce for all districts in India. Later on, this has been merged with `level_03.dta` and every state apart from Tamil Nadu has been trimmed for Gini Calculation, and limiting ourselves to Tamil Nadu only.
  - `final`: This contains the final dta file, which has districts of Tamil Nadu, and corresponding MPCE Gini Coefficients.
- `03. Code`: This has all the codes that we have used for cleaning the HCES dataset, and generating Gini Coefficients and ELMO measure in districts of Tamil Nadu. Below is a brief description about every code file:
  - `00_master.do`: This do-file will set the root directories to all the do-files which you will be exploring. Please make sure you set your directory to the folder `Discrimination_Assignment_ABM` on your device, set a global macro `'dir'` (example given in `00_master` itself) and then run this do-file before you run any other do file.
  - `01_Clean.do`: This do-file consists of the cleaning codes of HCES raw microdata.
  - `02_MPCE.do`: This do-file consists of the codes that have been used to calculate the MPCE for all districts in India.
  - `03_gini_gen.do`: This do-file consists of the codes used for Gini Coefficient Calculation.
  - `04_Heatmaps.do`: This do-file consists of codes for fuzzy-matching in STATA, and codes for generating Gini Heatmaps.
  - `05_ELMO_calculation.do`: This do-file consists of the stata codes that have been used to calculate ELMO measures for religions and caste categories in Tamil Nadu.

  Apart from these STATA do files, we have used some Python codes as well. The description for the python notebooks are given below:
  - `01_kml_shp_convert.ipynb`: When the Tamil Nadu shapefile was downloaded, it was in .kml format. To make it user-friendly, and accessible on all platforms, we have converted it to .shp. The codes for this conversion can be found on this notebook.
- `04. Latex`: This folder contains another folder named `Figures`, and this has the graphs which we have generated during this assignment. The important graphs have been attached in Appendix - 1 of this document.
- `05. Docs`: This folder consists of the HCES data documentations.

References

Elbers, C., Lanjouw, P., Mistiaen, J. A., & Özler, B. (2008). Reinterpreting between-group inequality. *The Journal of Economic Inequality*, *6*, 231–245.