Assessing Sensitivity of Income Inequality Measurements to Gerrymandering

Alexander Long, Megan Luick
Research Question

How are income inequality measurements affected by the administrative units used to spatially aggregate the data?

Medium income in MSP, calculated on county subdivisions vs. census tracts.
Problem Statement

Input:
- Geolocated income records
- Administratively-defined hierarchical spatial partitions
- Existing income inequality measures

Output:
- Inequality calculations on existing and randomly-defined spatial partitions

Objective:
- Describe the sensitivity of inequality measurements to changes in partition shape and scale

Constraints:
- Random partitions subject to equal area / equal population constraints
Societal Motivation

- Income inequality is one of the major problems of our time
- Poverty is linked to lower life expectancy, mental health disorders
- Economic inequality leads to distrust in political leadership
- An accurate portrait of inequality can improve policy and lives

‘Parasite’ paints a nightmarish picture of Korean inequality. The reality in America is even worse.
Societal Motivation

- Policy makers measure inequality at varying spatial resolution
  - Countries - cities - neighborhoods
- We cannot rely on inequality measures to be insensitive to:
  - Scale
  - Spatial partitioning
- Often, only aggregate data is publicly available

How can we inform public policy from a spatial perspective?
Key Concepts

- **Modifiable Areal Unit Problem (MAUP).** Results of spatial computation vary with zone.
- When calculating multivariate statistics and varying zone size and shape, no general trend exists.
  - “Not a well behaved problem"

Gerrymandering: Figure by Steve Nass (Wikimedia Commons)
Key Concepts

- **US Census Data.** Publicly-available down to the block level
- Within spatial partitions, means and margins of error are reported
- Microdata (individual records) available for Public Use Microdata Areas (PUMAs); not geolocated
- Complete microdata is made public... after 72 years
Key Concepts

- **Income Inequality Statistics.** Many exist, measuring “deprivation” or “variation” in income.
- The choice of statistic tends to matter more than choice of zone.
- Gini Coefficient is commonly used but is sensitive to spatial aggregation
  - Population-weighted and neighbor-weighted (“spatial”) Gini have been proposed.

<table>
<thead>
<tr>
<th>Coefficient of variation (CV) (unweighted)</th>
<th>Population weighted coefficient of variation (Williamson index (WI))</th>
</tr>
</thead>
<tbody>
<tr>
<td>[ CV = \frac{1}{n} \sum_{i=1}^{n} (y_i - \bar{y})^2 ]</td>
<td>[ WI = \frac{1}{\bar{y}} \sum_{i=1}^{n} (y_i - \bar{y})^2 ]</td>
</tr>
<tr>
<td>Theil index (TE(0))</td>
<td>Atkinson index (AT)</td>
</tr>
<tr>
<td>[ TE(0) = \frac{1}{n} \sum_{i=1}^{n} \log \frac{y_i}{\bar{y}} ]</td>
<td>[ AT = 1 - \frac{1}{\bar{y}} \sum_{i=1}^{n} (y_i - \bar{y})^1 - \bar{y} ]</td>
</tr>
<tr>
<td>Hoover coefficient (HC)</td>
<td>Coulter coefficient (CC)</td>
</tr>
<tr>
<td>[ HC = \frac{1}{\bar{y}} \sum_{i=1}^{n} \frac{A_i}{A_{tot}} (y_i - \bar{y}) ]</td>
<td>[ CC = \frac{1}{\bar{y}} \sum_{i=1}^{n} (\frac{A_i}{A_{tot}} y_i - \bar{y} \frac{A_i}{A_{tot}})^2 ]</td>
</tr>
<tr>
<td>Gini (U) (unweighted)</td>
<td>Gini (W) (population weighted)</td>
</tr>
<tr>
<td>[ Gini = \frac{1}{2n^2} \sum_{i=1}^{n} \sum_{j=i+1}^{n}</td>
<td>y_i - y_j</td>
</tr>
</tbody>
</table>

*From Portnov & Felsensetin (2010)*

*Spatial Gini*” proposed by Rey & Smith (2012)
Limitations of Existing Work

- Acknowledges MAUP effect?
  - Yes
    - Quantifies sensitivity of empirical measurements?
      - Yes
        - Addresses income inequality?
          - Yes
            - Our work
          - No
            - No
              - Portnov & Felsenstein (2010)
              - John Logan (2010)
              - Rey & Smith (2013)
      - No
        - Gini (1912)

- Yes
  - Quantifies sensitivity of empirical measurements?
    - Yes
      - Addresses income inequality?
        - Yes
          - Putnam & Chung (1989)
          - Fotheringham & Wong (1991)
          - Amrhein & Flowerdew (1992)
          - Briant et al. (2010)
        - No
          - No
How can we solve it?

- Use microdata rather than aggregates
- Experiment:
  - Rank inequality measurements by how sensitive they are to scale and choice of zone
  - Understand the extent at which bias exists in US census zones
Experimental Methodology

Dataset:

1) Synthetic data, based on 2010 income distribution, location randomly assigned within Census block groups.
2) Real locations from 1880 census, income data based on median occupation income.
Experimental Methodology

Computations to Compare

- Calculating many common income inequality measurements:
  - Theil Index
  - Hoover Coefficient
  - Williamson Index
  - Traditional Gini
  - Population-weighted
  - Spatial Gini

<p>| Table 1 |</p>
<table>
<thead>
<tr>
<th>Commonly used measurements of regional inequality.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coefficient of variation (CV) (unweighted)</td>
</tr>
<tr>
<td>[ CV = \frac{1}{\bar{y}} \sum_{i=1}^{n} (y_i - \bar{y})^2 ]</td>
</tr>
<tr>
<td>Theil index (TE(0))</td>
</tr>
<tr>
<td>[ TE(0) = \frac{1}{n} \sum_{i=1}^{n} \log \frac{y_i}{\bar{y}} ]</td>
</tr>
<tr>
<td>Hoover coefficient (HC)</td>
</tr>
<tr>
<td>[ HC = \frac{1}{2} \sum_{i=1}^{n} \left( \frac{A_i}{A_{tot}} \right) (y_i \bar{y} - \frac{A_i}{A_{tot}}) ]</td>
</tr>
<tr>
<td>Gini (U) (unweighted)</td>
</tr>
<tr>
<td>[ Gini = \frac{1}{2 \bar{y}} \sum_{i=1}^{n} \sum_{j=1}^{n} \left</td>
</tr>
<tr>
<td>Population weighted coefficient of variation (Williamson index (WI))</td>
</tr>
<tr>
<td>[ WI = \frac{1}{\bar{y}^2} \sum_{i=1}^{n} (y_i - \bar{y})^2 \left( \frac{A_i}{A_{tot}} \right)^{1/2} ]</td>
</tr>
<tr>
<td>Atkinson index (AT)</td>
</tr>
<tr>
<td>[ AT = 1 - \frac{1}{\bar{y}^{2(1-\epsilon)}} \sum_{i=1}^{n} (y_i \bar{y}^{1-\epsilon} - \frac{A_i}{A_{tot}})^{1/(1-\epsilon)} ]</td>
</tr>
<tr>
<td>Coulter coefficient (CC)</td>
</tr>
<tr>
<td>[ CC = \frac{1}{2} \sum_{i=1}^{n} \left( \frac{A_i}{A_{tot}} \right) (y_i \bar{y} - \frac{A_i}{A_{tot}})^{2/1} ]</td>
</tr>
<tr>
<td>Gini (W) (population weighted)</td>
</tr>
<tr>
<td>[ Gini = \frac{1}{2 \bar{y}} \sum_{i=1}^{n} \sum_{j=1}^{n} \left( \frac{A_i}{A_{tot}} \right) (y_i \bar{y} - \frac{A_i}{A_{tot}}) \left</td>
</tr>
</tbody>
</table>

From Portnov & Felsensetin (2010)
Experimental Methodology

Experiment Design

- Monte Carlo (bootstrapping) experiment
- "Sample" many possible administrative zones according to these criteria:
  - Comparable areas
  - Comparable populations
  - Comparable densities
- Hypothesis test 1: inequality measurement on census tracts / blocks / block groups are expected from the generated distribution of zones.
- Hypothesis test 2: inequality measurement is sensitive to choice of spatial partitioning - Gini underestimates at higher levels of aggregation.
Timeline

- 3/9: Datasets obtained
- 3/24: Able to generate a random collection of spatial zones given a criteria. Paper rough draft complete by week's end.
- 3/31: Able to generate all inequality statistics for a given zone
- 4/7: Preliminary analysis and conclusions from our experiments written into rough draft
- 4/14: Paper revisions due
- 4/21: Final presentations begin
Questions?

Or, email feedback: longx552@umn.edu
luick006@umn.edu