Measuring inequality

LECTURE 13
Outline for final lectures

- Once datasets have been finalized, it is time to produce results, with the aim of representing the patterns emerging from the data.

- In practice?
  - Inequality  
  - Poverty  
  - Basic summary statistics on household demographics, education, access to services, etc.
  - Average expenditures and incomes

this lecture  
next lecture  
final lecture
Inequality and poverty measurement

1) a measure of living standards
2) high-quality data on households’ living standards
3) a distribution of living standards (inequality)
4) a critical level (a poverty line) below which individuals are classified as “poor”
5) one or more poverty measures
Cowell (2011)

99.9% of this lecture is explained with better words in Cowell’s work: this book and other (countless) journal articles.
Warning

- During the course we paid attention to distinguish between different concepts: living standard, income, expenditure, consumption, etc.

- In this lecture we make an exception and use these terms interchangeably – we focus on measuring inequality of “a distribution”

- Similarly, I will not make a distinction between income per household, per capita, or per adult equivalent

- For once, and for today only, we will be (occasionally) inconsistent
Basic concepts

- Economists make a distinction between:
  - **Functional** distribution of income
    - distribution among factors of production
      land (rent), labor (wages), and capital (profits)
  - **Personal** (or size) distribution of income
    - distribution among persons, irrespective of their economic function

- We focus on the latter.
Functional Distribution of National Income in India

Rahul H. Dholakia

Apart from providing insights into the relationship between economic development and functional distribution of national income, estimates of factor shares in a developing country serve several other useful purposes. This article presents the functional distribution of national income in India over the period 1960-61 to 1991-92. An attempt is made to estimate the factor share of labour, land, and capital (including enterprises) not only for the economy as a whole but also for the broad categories of primary, secondary and tertiary sectors as well as the public and private sectors.

Introduction

Classical as well as neo-classical economists have shown that the measures of income like national income, national product and total value of the goods and services produced, etc., provide incomplete information about the economic phenomena. In particular, the concept of income attributed to land, capital and labour are not measurable in a way that ensures that the sum of these factor incomes is equal to the national income. Economists have therefore added another concept of income, the national income measure: the sum of the income earned by, and the income attributed to, the factors of production. The shares of income thus attributed to the factors of production are important for the analysis of income distribution and the study of economic development. The shares of income attributed to the factors of production are important for the analysis of income distribution and the study of economic development as the factors of production are the primary factors in the modern economic system.

Basic Data and Methods of Estimation

The data presented in Table 1, which has been extracted from other tables, shows the distribution of income among the three factors of production. The data are presented for the economy as a whole and for the primary, secondary and tertiary sectors.

Table 1: Distribution of Income in India, 1960-61 to 1991-92

<table>
<thead>
<tr>
<th>Sector</th>
<th>Labour</th>
<th>Land</th>
<th>Capital</th>
</tr>
</thead>
<tbody>
<tr>
<td>Primary sector</td>
<td>56.42</td>
<td>30.30</td>
<td>13.28</td>
</tr>
<tr>
<td>Secondary sector</td>
<td>67.68</td>
<td>3.47</td>
<td>28.85</td>
</tr>
<tr>
<td>Tertiary sector</td>
<td>61.57</td>
<td>3.74</td>
<td>34.69</td>
</tr>
<tr>
<td>All sectors</td>
<td>60.69</td>
<td>15.21</td>
<td>24.10</td>
</tr>
<tr>
<td>Public sector</td>
<td>86.15</td>
<td>0.83</td>
<td>13.02</td>
</tr>
<tr>
<td>Private sector</td>
<td>56.53</td>
<td>17.76</td>
<td>25.71</td>
</tr>
</tbody>
</table>

Dholakia (1996), Functional Distribution of National Income in India, Economic and Political Weekly
Focus on the term 'inequality'

- “When we say income inequality, we mean simply differences in income, without regard to their desirability as a system of reward or undesirability as a scheme running counter to some ideal of equality” (Kuznets 1953: xxvii)

- In practice, how can we appraise the inequality of a given income distribution? Three main options:
  1. Tables
  2. Graphs
  3. Summary statistics
The distribution of income and taxable income
2018 Tax statistics, National Treasury and the South African Revenue Service
<table>
<thead>
<tr>
<th>Tax year</th>
<th>2017</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Number of taxpayers</td>
</tr>
<tr>
<td>&lt;= 0</td>
<td>116 998</td>
</tr>
<tr>
<td>1 – 70 000</td>
<td>401 447</td>
</tr>
<tr>
<td>70 001 – 350 000</td>
<td>2 689 263</td>
</tr>
<tr>
<td>350 001 – 500 000</td>
<td>764 197</td>
</tr>
<tr>
<td>500 000 +</td>
<td>926 660</td>
</tr>
<tr>
<td>Total</td>
<td>4 898 565</td>
</tr>
</tbody>
</table>
Tables: an assessment

- In general, tables are **not recommended** when the focus is inequality.
- Difficult to get a clue of the extent of inequality in the distribution by looking at a table, plus income brackets are arbitrary.
- Does putting income distribution into a graph (diagram) **help** to represent inequality?
Histograms

- Let the interval $[x^-, x^+]$ denote the range of the data.

- Partition $[x^-, x^+]$ into $m^*$ non-overlapping bins (intervals) of equal width $h = (x^+ - x^-)/m^*$.

- A histogram estimate of the density $f(x)$ is the fraction of observations falling in the bin containing $x$, divided by the bin width $h$:

$$\hat{f}(x) = \frac{\text{(fraction of sample obs. in same bin as x)}}{h}$$

- The area of each bar ($= h \times \hat{f}(x)$) is interpreted as the fraction of sample observations within the bin. All bar areas sum up to unity.
Histograms?
Histograms: an assessment

- The position and number of bins is *arbitrary*
- Inherently *lumpy*: discontinuities at the edge of each bin
- Can provide very *different pictures* of the same distribution
- Read Cowell, Jenkins and Litchfield (1996) for more.
Beyond histograms

- A **probability density function (PDF)** is the ‘continuous version’ of a histogram

- A convenient way to introduce the PDF is by starting from the **cumulative distribution function (CDF)**
The cumulative distribution function (CDF)

- The cumulative distribution function (CDF) is defined as follows:

\[ F(x) = \int_{0}^{x} f(x) \, dx \]

- \( F(x) \) is the proportion of individuals having \( X \) less than or equal to \( x \).

- If \( X \) is income and, say, \( x = 2,000 \) Rps., then \( F(x) = \Pr(X < 2,000) \), that is the fraction of people with less than 2,000 Rps.
The empirical cumulative distribution function (ecdf)

Mongolia HSES 2016, Cumulative distribution of per capita consumption (p.10)

- Pick up any income level on the $x$-axis, and the curve $F(x)$ will tell you the percentage of individuals in the population having a level of income lower than $x$. 

Source: HSES 2016
The probability density function (pdf)

- The probability distribution function (pdf) is the derivative of the CDF:

\[
f(x) = \frac{dF(x)}{dx}
\]

- By definition of derivative:

\[
f(x) = \lim_{h \to 0} \frac{F(x + h) - F(x)}{h}
\]
The probability density function (pdf)

- Now drop the limit (and replace = by ≈):

\[
f(x) \approx \frac{F(x + h) - F(x)}{h}
\]

- The PDF \( f(x) \) is not a probability measure, but a scaled version of it: it is the probability of \( X \) falling in the interval \((x, x + h)\) divided by the length \( h \) of such an interval.
Mongolia, 2016

PDF of per capita consumption (p.11)
Interpretation

Two extremes:

a) **perfect equality**: everyone is concentrated at one particular income value

b) **uniform density**: income is spread uniformly from the poorest to the richest individual – significant inequality

c) **in-between, typical case**.
Pdfs: an assessment

- The bandwidth is arbitrary

- In most cases, they require some **trimming** of top values to avoid looking “squished” and being unreadable

- In general, it does not show what is going on in the **upper tail** very clearly
The quantile function

- Let $p = F(x)$ be the proportion of people in the population with income lower than $x$.
- The quantile function $Q(p)$ is defined as:
  $$F\left[Q(p)\right] = p \quad \text{or} \quad Q(p) = F^{-1}(p)$$
- $Q(p)$ is the income level below which we find a proportion $p$ of the population.


*Pen’s Parade (Quantile Function) for Expenditure per Capita, Vietnam, 1998*
The Lorenz Curve (1905)

Picture and intuition

- **Horizontal axis:** cumulative % of population (individuals ordered poorest to the richest)
- **Vertical axis:** cumulative % of income received by each cumulative % of population.
- **45-degree line:** Lorenz curve if perfect equality.
- The overall distance between the 45-degree line and the Lorenz curve is indicative of the amount of inequality present in the population.
The Lorenz Curve (1905)
Mathematically

- The Lorenz curve $L(p)$ is defined as follows:

$$L(p) = \frac{\int_0^p Q(p) \, dq}{\int_0^1 Q(p) \, dq}$$

- The numerator sums the incomes of the poorest $p\%$ of the population;
- The denominator sums the incomes of all.
- The ratio $L(p)$ indicates the cumulative $\%$ of total income held by a cumulative proportion $p$ of the population.
- Example: if $L(0.5) = 0.3$, then we know that the $50\%$ poorest individuals hold $30\%$ of the total income in the population.
Quantile function and Lorenz curve: an assessment

- These graphical tools emphasize the *ranking* of shares of the population on the basis of income
- The Lorenz curve clearly shows how far the distribution is from perfect equality
- Still, no graph is as straightforward and easily comparable as a *scalar measure* of inequality
Recap and next steps

- Not all graphs are OK to represent inequality

- **Lorenz curve** is the most popular

- A better conceptual understanding comes from constructing inequality measures from first principles.

- The most straightforward approach: inequality measures as pure statistical measures of dispersion.
Inequality indicators
Measures of dispersion

- **Range** \( R = x_{\text{max}} - x_{\text{min}} \)

  ▲ **PRO**: Easy to compute and communicate
  ▼ **CON**: Insensitive to changes between extremes (can we really know min and max?)

- **Variance** \( \sigma^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \mu)^2 \)

  ▲ **PRO**: Easy to compute, additively decomposable
  ▼ **CON**: not robust (outliers), depends on the scale of measurement

- **Coefficient of Variation** \( CV = \frac{\sqrt{\sigma^2}}{\mu} \)

  ▲ **PRO**: Scale invariant
  ▼ **CON**: not robust (outliers), properties?
The p-quantile of a distribution of values is a number $x_p$ such that a proportion $p$ of the population values are less than or equal to $x_p$.

For example, if $p = 0.5$, then the 0.5-quantile $x_{0.5}$ is any value such that $F(X < x_{0.5}) = 0.5$.

Certain quantiles have special names:

- The 0.5-quantile $x_{0.5}$ is the median, or 50-th percentile.
- The 0.1-quantile is the first decile, or 10-th percentile.
- The 0.2-quantile is the first quintile, or 20-th percentile.
- The 0.25-quantile is the first quartile $Q_1$, or 25-th percentile.
- etc. etc.
Quantile ratios

- A quantile ratio measures the gap between the rich and the poor.
- It is defined as the ratio of two quantiles, $Q(p_2)/Q(p_1)$, using percentiles $p_1$ and $p_2$.

Three popular indices are:

- The quintile ratio ($p_2 = 80$ and $p_1=20$):
  $$QR = Q(p_{80})/Q(p_{20})$$

- The decile ratio ($p_2 = 90$ and $p_1=10$):
  $$DR = Q(p_{90})/Q(p_{10})$$
The decile ratio
per adult equivalent income (OECD def.)

25.6

[Bar chart showing the decile ratio for various countries, with Denmark having the highest ratio at 25.6.]
Quantile share ratios

- Let $S_{20}$ denote the share of (equivalised disposable) income received by the bottom 20% of the population, and $S_{80}$ the income share received by the top 20% of the population.

- The quintile share ratio is defined as follows:
  \[ S_{80-20} = \frac{S_{80}}{S_{20}} \]

- The quintile share ratio is the level-1 Laeken indicator, chosen by the EU to monitor income distribution.
QSR around the world

Source: (WDI) Income share held by highest 20% over income share held by lowest 20%, last available 2010-2017
$S_{80}/S_{20}$ in Sub Saharan Africa

Source: (WDI) Income share held by highest 20% over Income share held by lowest 20%, last available 2010-2017
The Gini Coefficient

A definition

- Yitzhaki (1997) counts more than a dozen formulas available for the Gini index.

- A classic definition of the Gini coefficient:

\[ G = \frac{1}{2n^2 \mu} \sum_{i=1}^{n} \sum_{j=1}^{n} |x_i - x_j| \]

- The Gini coefficient ranges from 0 (all recipients have the same income: full equality), to 100 (all income is received by one recipient: maximum inequality).
The Gini Coefficient
Interpretation – Pyatt 1976: 244

- The Gini index “is the average gain to be expected, if each individual has the choice of being himself or some other member of the population drawn at random, expressed as a proportion of the average level of income”

- E.g., if the Gini index for an Italian is 0.30, we can say that the expected gain from playing the experiment of exchanging income with someone else randomly chosen in the Italian population, is 30% of average income.
The Gini Coefficient

**graphical interpretation**

- The Gini index is **two times the area** $A$ **between the Lorenz curve and the equality diagonal:**

\[
Gini = \frac{A}{(A + B)}
\]

\[
= 2A
\]

\[
= 2 \left( \frac{1}{2} - B \right) = 1 - 2B
\]

\[
Gini = 2 \int_{0}^{1} [p - L(p)] dp
\]

\[
= 1 - 2 \int_{0}^{1} L(p) dp
\]
Gini index around the world

Source: (WDI), GINI index (World Bank estimate) last available year 2010-2017
The Gini index in Sub-Saharan Africa

Source: (WDI), GINI index (World Bank estimate) last available year 2010-2017
Gini around the World
A selection of countries

Source: (WDI), GINI index (World Bank estimate) last available year 2010-2017
Recap

- Quantile ratios, quantile share ratios, Gini, are all popular inequality measures
- They do a fine job at representing inequality with a number
- **Problem**
  - they do not always have all the properties that we would want for an inequality measure
- **Solution**
  - solve the problem backwards. First lay out some desirable properties, then construct a measure that complies with them
Deriving inequality measures from axioms

- **Axiom**: a statement accepted as true as the basis for argument or inference.

- The **axiomatic approach** allows us to “custom-build” inequality measures that fit our needs:
  1. We define a set of elementary properties (axioms) that we think inequality measures ought to have
  2. We obtain a mathematical formula that delivers a class of inequality measures satisfying the axioms
A. **Anonymity (or Symmetry)**
   *Who* is earning the income does not matter

B. **The Population Principle**
   *Population size* does not matter

C. **Scale Invariance (or Relative Income Principle)**
   Income *levels* do not matter

D. **The (Pigou-Dalton) Principle of Transfers**
   Rank-preserving rich-to-poor transfers reduce inequality

E. **Decomposability (or Subgroup Consistency)**
   The measure is additively decomposable
Generalized Entropy Indices (GEI)
Shorrocks (1980)

- Inequality measures that satisfy all axioms (A to E), must have the following form:

\[ GE(\theta) = \frac{1}{\theta^2 - \theta} \left[ \frac{1}{n} \sum_{i=1}^{n} \left( \frac{x_i}{\bar{x}} \right)^\theta - 1 \right] \]

where \( \theta \) is a parameter that may be given any value (positive, zero or negative).
The Generalized Entropy Indices

- Depending on the value of the $\theta$ parameter:
  
  $\theta = 0 \Rightarrow$ Mean Logarithmic Deviation
  
  $GE(0) = MLD = \frac{1}{n} \sum_{i=1}^{n} \log \left( \frac{\bar{x}}{x_i} \right)$

  $\theta = 1 \Rightarrow$ Theil Index
  
  $GE(1) = THEIL = \frac{1}{n} \sum_{i=1}^{n} \frac{x_i}{\bar{x}} \log \left( \frac{x_i}{\bar{x}} \right)$

  $\theta = 2 \Rightarrow$ Half Coefficient of Variation Squared
  
  $GE(2) = \frac{\kappa^2}{2}$
Inequality decompositions are typically used to estimate the extent to which the **heterogeneity** of the population affects overall inequality. Two popular techniques are:

1. Decomposition by population sub-group
2. Decomposition by income source

We focus on the former:

Societies can often be partitioned into groups (e.g. North-South). We would like to be able to decompose total inequality into two components, namely the inequality **within** the constituent groups, and inequality **between** the groups:

\[ I_{TOTAL} = I_{WITHIN} + I_{BETWEEN} \]
Inequality decomposition

- The most popular additively decomposable inequality index is the Mean Logarithmic Deviation.

- Partition the population into \( k = 1, \ldots, K \) groups. Then:

\[
MLD = \sum_{k=1}^{K} \nu_k MLD_k + \sum_{k=1}^{K} \nu_k \log \left( \frac{\bar{x}}{\bar{x}_k} \right)
\]

\( \nu_k \) are population shares.
## Botswana, 2009/10

household income and expenditure survey

<table>
<thead>
<tr>
<th></th>
<th>Total</th>
<th>2010</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>GE(0)</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td></td>
<td>0.669</td>
</tr>
<tr>
<td><strong>Urban / rural</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Between-group</td>
<td></td>
<td>0.063</td>
</tr>
<tr>
<td>inequality</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Between as a share</td>
<td></td>
<td>0.094</td>
</tr>
<tr>
<td>of total</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Within-group</td>
<td></td>
<td>0.606</td>
</tr>
<tr>
<td>inequality</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Region</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Between-group</td>
<td></td>
<td>0.060</td>
</tr>
<tr>
<td>inequality</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Between as a share</td>
<td></td>
<td>0.090</td>
</tr>
<tr>
<td>of total</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Within-group</td>
<td></td>
<td>0.609</td>
</tr>
<tr>
<td>inequality</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Lessons learned

- Many ways to describe inequality, some more effective than others

- **Graphs**: most notable are quantile functions and Lorenz curves

- **Measures**: different inequality measures lead to different results. Based on their properties, the recommended choice is GEI (generalized entropy indices), and in particular the MLD (mean log deviation). However, Gini remains extremely popular in practice.
References

**Required readings**


**Suggested readings**


Thank you for your attention
Homework
Exercise 1 – Engaging with the literature

Considering equations (10) to (12) in Farris (2010) give a brief interpretation of a Gini index of 63% for South Africa.
Exercise 2 - Inequality in South Asia

- Turn to page 2 of this report (see next slide)
- What criticisms would you make to this chart?
FIGURE 1  Based on standard monetary indicators, South Asia has moderate levels of inequality

Note: Orange and light brown bars indicate countries where inequality is estimated based on consumption per capita. Light blue bars indicate countries with estimates based on income per capita.
Exercise 3 - Functional vs Personal distribution of income

- The nature of the relationship that links the evolution of income shares to income inequality is complex and still widely debated among researchers.

- In that context, comment on Figure 19 of the ILO Global Wage Report 2016/2017.