TOWARDS A MORE REALISTIC ESTIMATE OF THE INCOME DISTRIBUTION IN MEXICO

Alfredo Bustos
INEGI, Mexico

Expert Meeting on Measuring Poverty and Inequality
Budva, Montenegro, 26-27 September 2017
alfredo.bustos@inegi.org.mx
LAYOUT

• Preliminaries
  – Problem
  – Adjustment to National Accounts (MSNA)

• Fitting Income Models
  – Purpose
  – Criterion
  – Models
  – Restrictions

• Numerical results.
Problem

• **Income**: important input for measurement of both poverty and inequality.

• However, not accurately measured through household income surveys.
Figure 13 - Saving as a percentage of adjusted disposable income by Equivalized Disposable Income quintile

Total Current Income in Mexico, 2012

\[2.43 \times \text{ENIGH} = \text{MSNA}\]
WHY?

• Survey income affected mainly by:
  – **Under-reporting**, and
  – **Truncation**: Households with very large income, absent from sample.

• By how much each affect total income?
  Not known
  – 100-0% or 0-100%

• Lacking evidence, consider both.
Attempt to account for the above difference

ADJUSTMENT TO NATIONAL ACCOUNTS
Adjustment to National Accounts

Survey: \( Y_{(i)} \), \( i = 1, \ldots, n \);  

\( \pi_{(i)} \)

\[ Y^{(a)}_{(i)} = f \left( Y_{(i)}, \hat{Y}_{SNA} - \hat{Y}_{Survey} \right), i = 1, \ldots, n, \]

So that:  
\[ \sum_{i=1}^{n} \frac{Y^{(a)}_{(i)}}{\pi_{(i)}} = \hat{Y}_{SNA}. \]

Income Distribution

Poverty, inequality and fiscal studies derived.
IGNORING TRUNCATION

![Graph showing percentage of population against log income.](attachment:graph.png)
IGNORING TRUNCATION

<table>
<thead>
<tr>
<th></th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>MSNA</td>
<td>$2,924,212,500,000</td>
</tr>
<tr>
<td>ENIGH</td>
<td>$1,203,201,383,751</td>
</tr>
<tr>
<td>HOUSEHOLDS (2012)</td>
<td>31,559,379</td>
</tr>
<tr>
<td>AVERAGE DIFFERENCE</td>
<td>$54,532.48</td>
</tr>
</tbody>
</table>

Poverty LOG INCOME ignoring truncation.

50%
ADD AVERAGE DIFFERENCE.

$$Y_{(i)}^{(a)} = Y_{(i)} + 54,532.48$$

- **MSNA**: $2,924,212,500,000
- **ENIGH**: $1,203,201,383,751
- **HOUSEHOLDS (2012)**: 31,559,379
- **AVERAGE DIFFERENCE**: $54,532.48

**IGNORING TRUNCATION**
IGNORING TRUNCATION

PROPORTIONAL ALLOCATION

\[ Y^{(a)}_i = Y^{(i)} \times 2.43 \]

(OCDE 2014)
(78%) 83% (88%) of difference added to the survey X-th income decile; (22%) 17% (12%), to IX-th decile. **Conclusion: inequality is extreme in Mexico.**

CONSEQUENCES

• In absence of evidence to support either correction of survey-declared incomes, **CONEVAL** (Mexican institution in charge of measuring poverty) decided **NOT TO CARRY OUT ANY INCOME CORRECTION**.

• Equivalent to assuming under-reporting and truncation absent.

• Risks: In presence of significant under-reporting, number of false positives (non-poor counted as poor) grows by unknown amount. On the other hand, inequality is underestimated.
FITTING MODELS TO SURVEY INCOME DATA
PURPOSE

• Skipping the intermediate step of imputing incomes in the sample, to estimate in a less arbitrary manner a household income distribution in Mexico which lies closer to reality by taking other data sources into consideration.
PROPOSED CRITERION
In order to reduce arbitrariness, criterion introduced to determine good and better. Some parametric distribution families fitted accordingly to quarterly current income. Also reduces arbitrariness. Use all available information.

Model: \( f(y; \theta) \Rightarrow \left\{ \ell(\theta; Y_{(i)}) = \ln(f(Y_{(i)}; \theta)); h(\theta) \right\} \)

Criterion: \( \operatorname{Max}_{\theta, \lambda} \left\{ \sum_{i=1}^{n} \frac{1}{\pi_{(i)}} \ell(\theta; Y_{(i)}) - \lambda'(h(\theta) - c) \right\} \)

ENIGH: \( \left\{ \frac{Y_{(i)}}{\pi_{(i)}}, i = 1, \ldots, n; \right\} \)

MSNA: \( c_1 = \text{Total}(\hat{Y}_{SNC}) \)

MIRS: \( c_2 = \text{Average}(Y_{\text{Max} - k}, \ldots, Y_{\text{Max}}) \)

## CONSTRAINTS

<table>
<thead>
<tr>
<th>Concepto</th>
<th>Restricción:</th>
<th>Interpretación</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average household income</td>
<td>$h_1(\theta) = E[Y \mid \theta] = c_1$</td>
<td>Mean income for fitted model equals average household income, according to MSNA.</td>
</tr>
<tr>
<td>Income Integral</td>
<td>$h_3(\theta) = \frac{1}{\alpha} \int_{\varphi_\alpha}^{\infty} y f_Y(y \mid \theta) dy = M = c_3$</td>
<td>Mean income for households whose income is greater than threshold $\varphi_\alpha$ is, according to the model, equal to average household income from SAT.</td>
</tr>
</tbody>
</table>
THRESHOLD DETERMINATION

• Requirements:
  – Informative of extreme conditions
  – Reduce weight of conceptual and of unit of observation differences.
• Explored combinations for $\alpha = 1\%, 0.1\%, 0.01\%, 0.001\% \text{ y } 0.0001\%$.
• First three under survey maximum.
• Last two fulfil our requirements.
NUMERICAL RESULTS
OPTIMAL FITTED MODELS

**Generalized Gamma (GG)**

- **OPTIMAL CMPL VALUE**: -365,349,780
- **GINI COEFFICIENT**: 0.62
- **RATIO X/I**: 52.09

**Type II Generalized Beta (GB2)**

- **OPTIMAL CMPL VALUE**: -365,541,743
- **GINI COEFFICIENT**: 0.61
- **RATIO X/I**: 51.66
LORENZ CURVES
INCOME DISTRIBUTION BY DECILES

ENIGH

GG

GB2

34.9%

53.8%

52.8%

20.5%

11.7%

12.0%
Gini Coefficients, Mexico, 2008-2014
Ratio X-th decile income to I-st decile, 2008-2014
Households with income below CONEVAL welfare lines, México, 2008-2014
Conclusion

• There is evidence of both income under-reporting and truncation in the sample.

• Best fitted model exhibits under-reporting growing with income but less than proportionally.

• Reality may be very different from survey results.
Conociendo México

01 800 11 44 24
www.inegi.org.mx
atencion.usuarios@inegi.org.mx

@inegi_informa

INSTITUTO NACIONAL DE ESTADÍSTICA Y GEOGRAFÍA

GRAÇHAS

HVALA

INEGI Informa