Income inequality and immobility in Latin America:
How much do we really know?

Francisco H. G. Ferreira

III / LSE

Drawing on joint work with Facundo Alvaredo, François Bourguignon, Paolo Brunori, Nora Lustig and Guido Neidhöfer

Within the Latin America and Caribbean Inequality Review (LACIR)

UCL Social Data Conference / LIS 40th Anniversary Conference

Latin American and Caribbean Inequality Review

LACIR is an independent scholarly endeavour created with the aim of understanding why, despite major structural economic and social change, inequality in Latin America and the Caribbean persists at exceptionally high levels.

International Inequalities Institute at the London School of Economics and Political Science

Inter-American Development Bank

Institute for Fiscal Studies

Yale University

THEMES

We study inequality in the region through five broad themes:

We hope that understanding the nature, causes and consequences of Latin America's stable high-inequality equilibrium may provide a basis for action intended to make the region more equitable.

Levels and trends of inequality

Establishing the facts about levels and trends of inequality in outcomes

VIEW THEME ONE

Inequality of opportunity

Analyzing the role of the family and communities in shaping inequality in outcomes and intergenerational mobility

VIEW THEME TWO

Inequality and markets

Studying the link between inequality and markets for labor, capital and goods

VIEW THEME THREE

Taxation and redistribution

Considering the limited role that fiscal redistribution plays in the region to level the playing field

VIEW THEME FOUR

Inequality and political power

Examining how inequality shapes political voice, political representation, social unrest and political outcomes

VIEW THEME FIVE

https://lacir.lse.ac.uk

Outline

“To know what you know and what you do not know: that is true knowledge”

- attributed to Confucius

1. Acknowledging and addressing uncertainty in the measurement of inequality and immobility in Latin America, given the prevailing data constellations. Two case studies:

   a) Levels and trends of income inequality: bands, not lines.

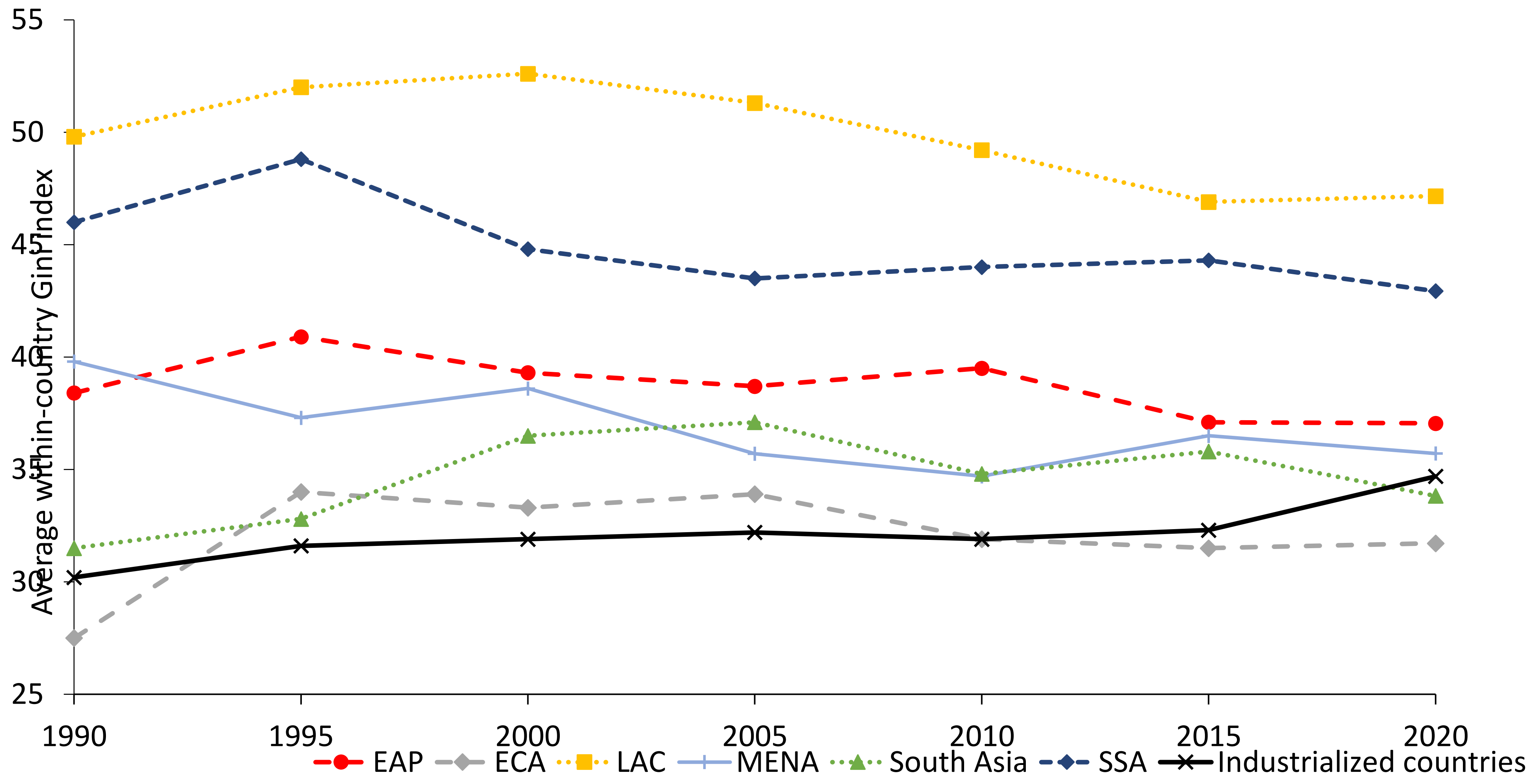
      - Drawing on Alvaredo, Bourguignon, Ferreira and Lustig: “Seventy years of income inequality in Latin America: Sources, methods and results”

   b) Inherited inequality: data and methods

      - Drawing on Brunori, Ferreira and Neidhöfer: “Inequality of opportunity and intergenerational persistence in Latin America”

1. Levels and trends of income inequality (in the World Bank data)

LAC in the global context over 30 years



Source: updates on Mahler, Yonzan and Lakner (2022), by courtesy of the authors.

Evidence from household surveys

- Those estimates come primarily from the PiP / POVCALNET database until 2015, but from a mix of sources for 2020 , including NSO estimates, phone surveys and projections based on NSA growth rates.
  - Data sources matter, and contain a range of different estimates

1. Some countries have more than one HH survey and survey methodologies differ.

2. Different authors and institutions make different adjustments to data, even from the same survey, e.g.:
   - Treatment of taxes and social security contributions
   - Equivalence scales
   - Trimming
   - Imputation of rents for owner-occupiers
   - Imputation of missing incomes

- Taken together, these can lead both to different concepts / definitions of income, and to different estimates of the same concept. See Székely & Hilgert (2007) for a pioneering study.
- Do HHS tell a consistent story when a wider range of sources is considered?

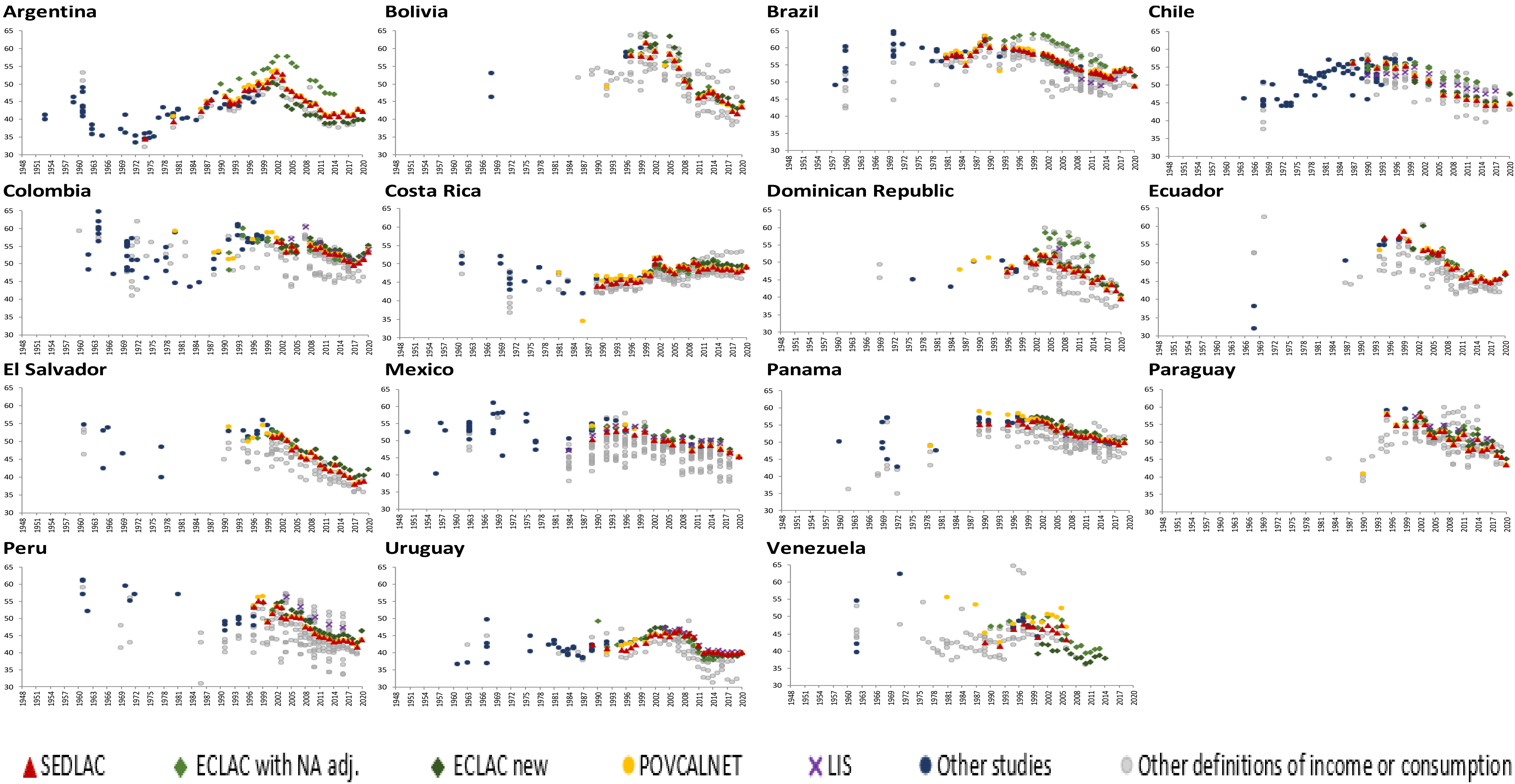
Evidence from household surveys

- To attempt to answer this question we have assembled a wide range of estimates for:
  - The Gini coefficient
  - Top 10% income shares (not shown today)
  - Middle 40% income shares (not shown today)
  - Bottom 50% income shares (not shown today)
- For the Gini coefficient alone, the next two slides contain 22,500 observations from 85 different sources or data producers, covering 33 LAC countries.
- Main sources:
  - UNU-WIDER database (a repository of multiple series)
  - Authors' own collection of historical studies
- Period: 1948-2020

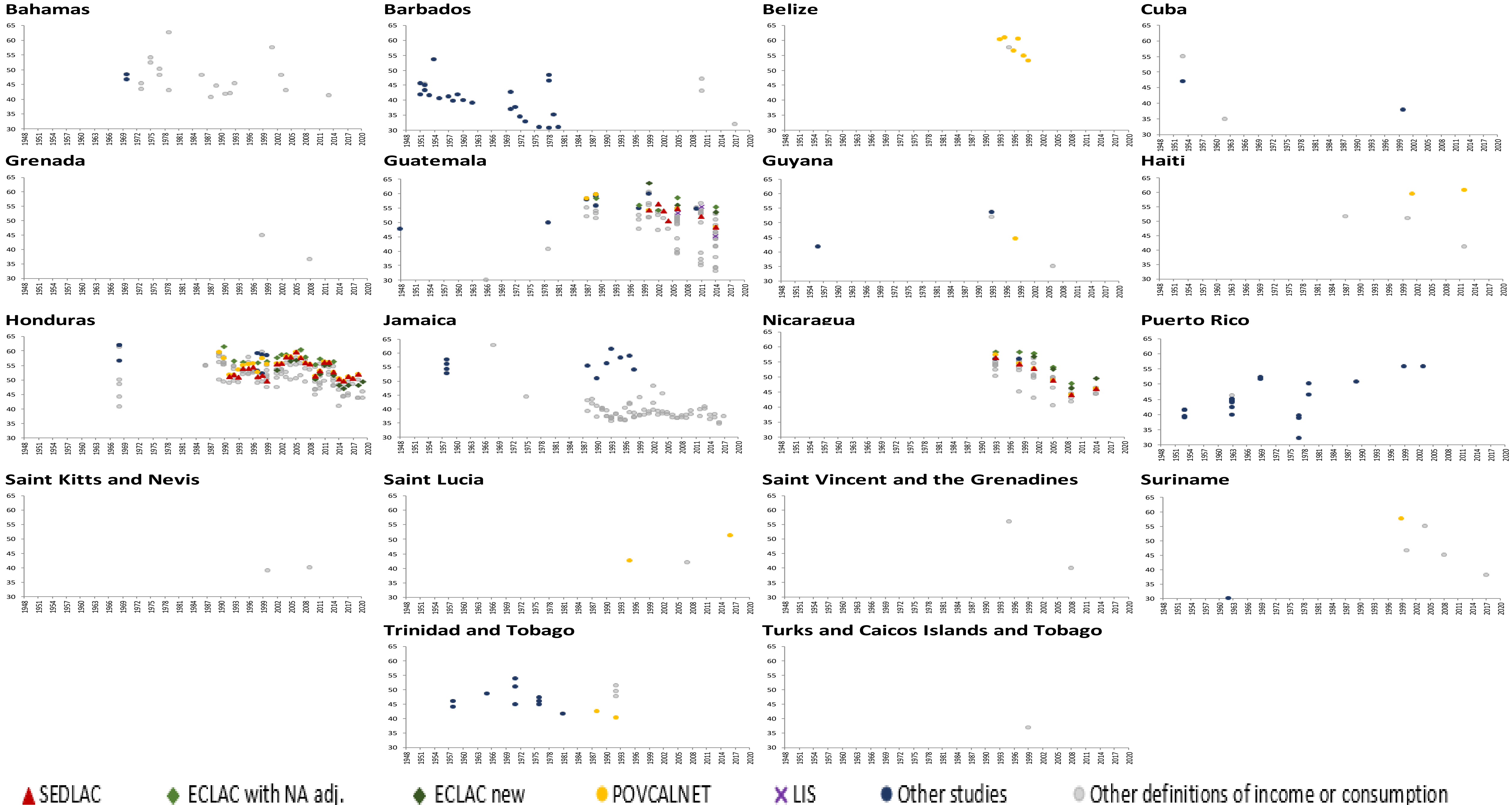
Evidence from household surveys

- The UNU-WIDER database reports series from a number of different sources, which rely on internally consistent methods and thus offer more comparable estimates. These include:
  - SEDLAC
  - POVCALNET (which is based on SEDLAC, and thus very similar, but not always identical)
  - CEPAL/ECLAC - “old” series: includes an adjustment based on National Accounts, following Altimir (1987)
  - CEPAL/ECLAC - new series: excludes that adjustment
  - LIS-Luxembourg Income Study, which has been progressively incorporated LATAM countries (namely Brazil, Chile, Colombia, Guatemala, Mexico, Panama, Peru, Paraguay and Uruguay)
- These series are highlighted in different colours. In addition:
  - Blue dots are individual observations from different studies which use household income per capita (or near enough) as income concept
  - Grey dots are observations that use other definitions of income (equivalized; total household income; consumption, earnings, etc.)

Gini coefficients: 15 countries, 1948-2020

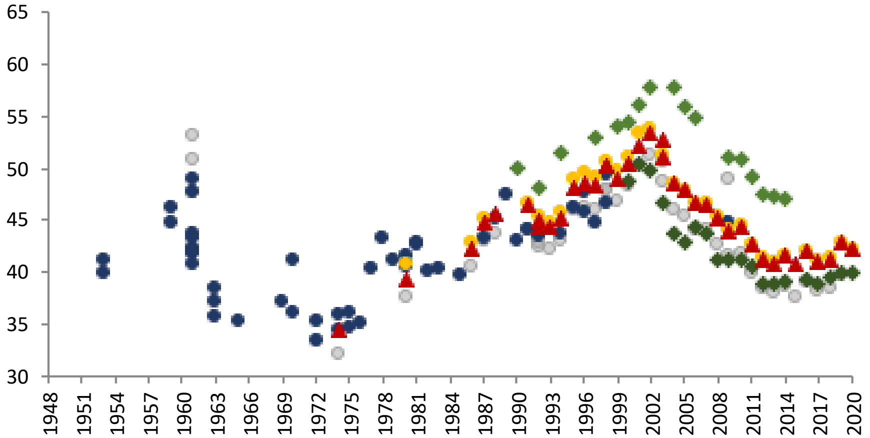


Gini coefficients: 18 countries, 1948-2020

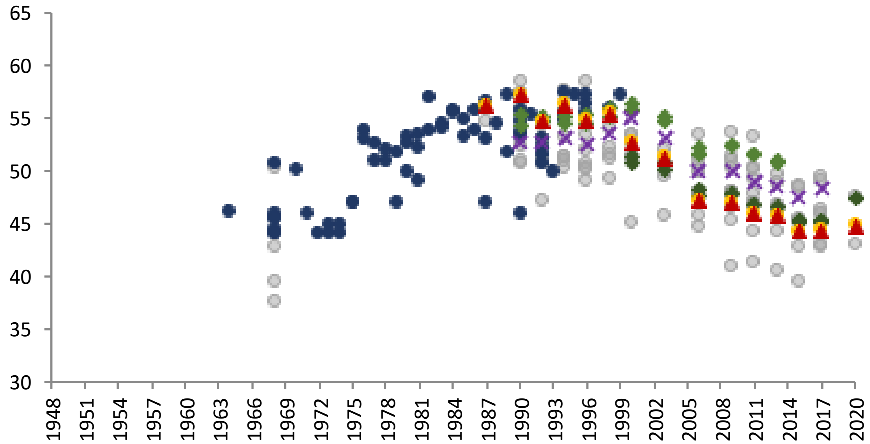


Evidence from household surveys: zooming in

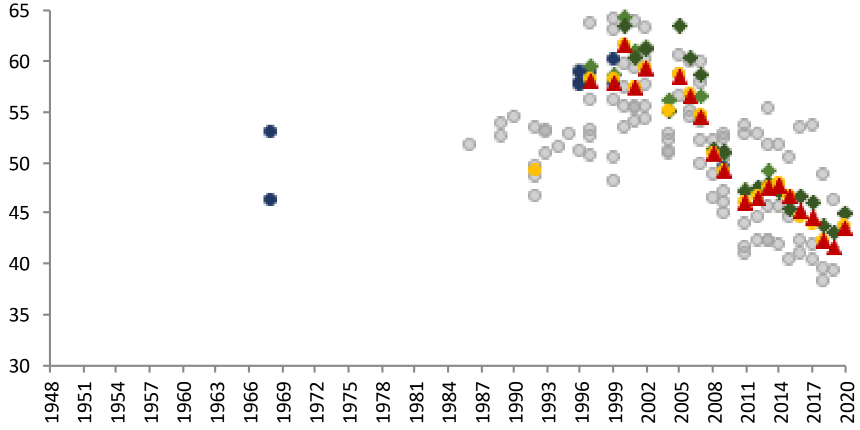
Argentina



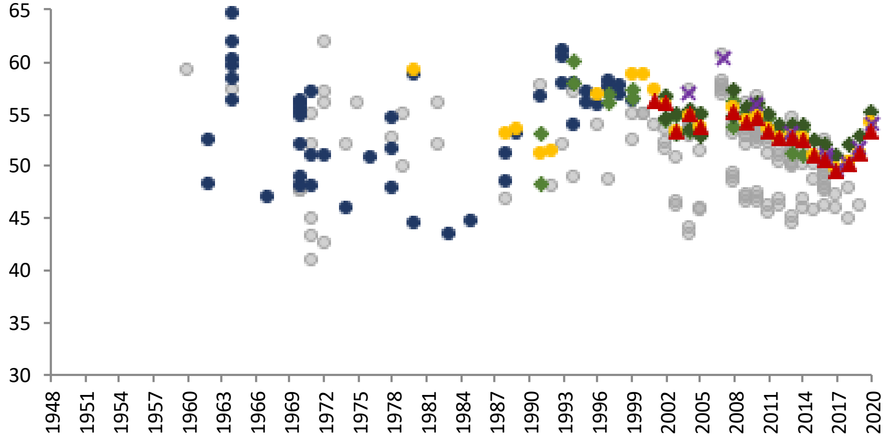
Chile



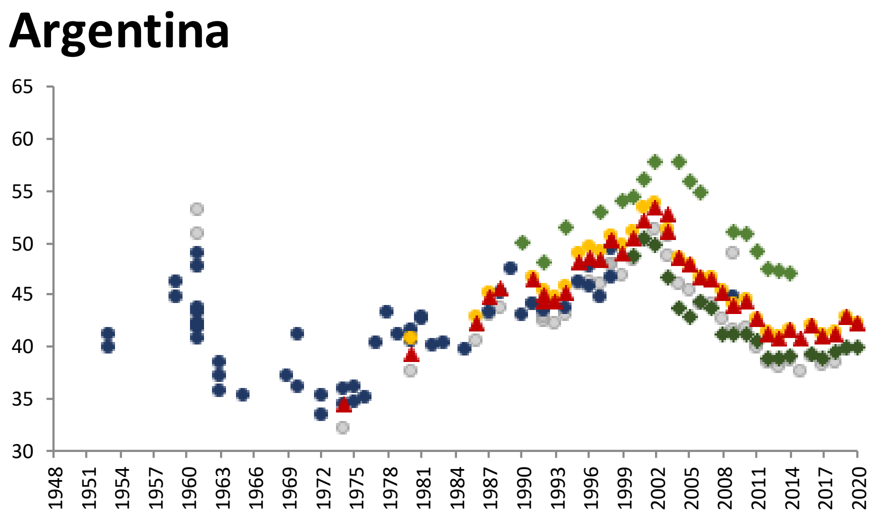
Bolivia

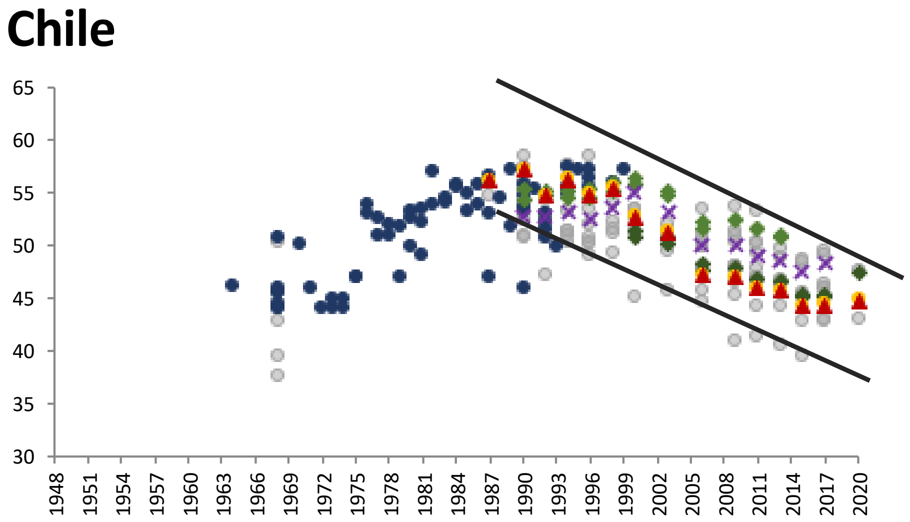


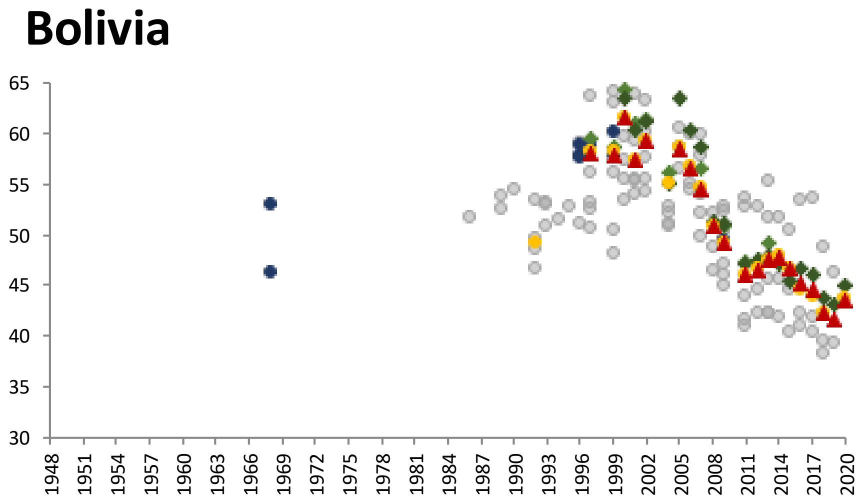
Colombia

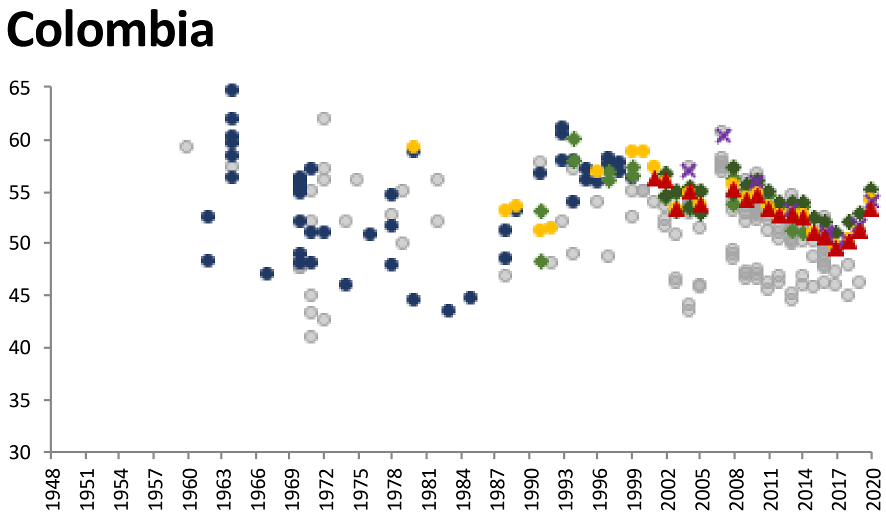


Evidence from household surveys: zooming in









Evidence from household surveys

1. There is considerable variation in **levels** for any given year.
    - It was particularly pronounced prior to the 1980s, when methodological differences also reflected attempts to "reconcile with SNA".
    - This variation is greater when other income concepts are included, but remains substantial even when only household income per capita is used.
    - The range is smaller when attention is restricted only to the 'harmonized series'. Even then, differences greater than 5 Gini points are not uncommon.
2. Inequality **trends** are much more robust, particularly since the 1980s
    - Different estimates form "inequality bands", which tend to evolve similarly.
3. This level uncertainty is before one introduces additional data sources, such as tax and social security records, or aggregate information from national accounts.

2. Inherited inequality: data and methods

- **Question:** How strong is the transmission of socioeconomic status across generations in Latin America?
  - Two main outcome variables used in the economics literature
    - Income
    - Education
    - (Occupation)
  - Two main approaches – both descriptive:
    - Intergenerational mobility (IGM)
    - Inequality of opportunity (IOp)

2. Inherited inequality: data and methods

Brief remark on the relationship between IGM and I.Op.

- IGM: How strongly are specific outcomes associated across generations?
- IOp: What share of current inequality can be accounted for by inherited (pre-determined) factors?
- These may seem quite different. Yet, for "origin-independence" mobility and one common measure of IOp:

$$y_c = \alpha + \beta y_p + \varepsilon \qquad \rho = \beta \frac{\sigma_p}{\sigma_c}$$

$$y_c = \alpha + C\beta + \varepsilon \qquad IOp^{ea} = \frac{I(\widehat{y_c})}{I(y)}$$

2. Inherited inequality: data and methods

Brief remark on the relationship between IGM and I.Op.

- IGM: How strongly are specific outcomes associated across generations?
- IOp: What share of current inequality can be accounted for by inherited (pre-determined) factors?
- These may seem quite different. Yet, for "origin-independence" mobility and one common measure of IOp:

$$y_c = \alpha + \beta y_p + \varepsilon \qquad \rho = \beta \frac{\sigma_p}{\sigma_c} \qquad \rho^2 = R^2$$

$$y_c = \alpha + C\beta + \varepsilon \qquad IOp^{ea} = \frac{I(\widehat{y_c})}{I(y)}$$

2. Inherited inequality: data and methods

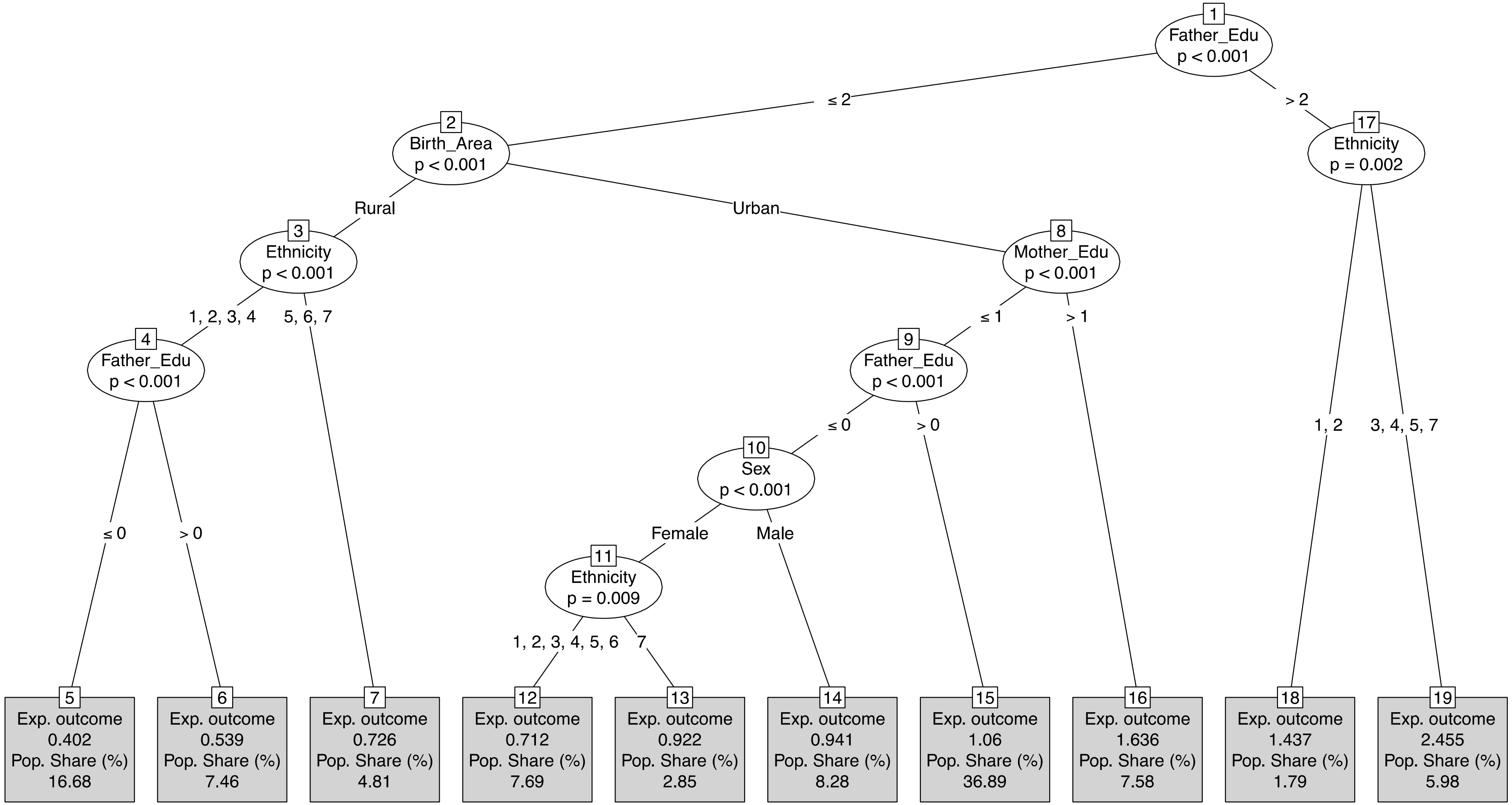
- Substantial literature estimating intergenerational mobility in both education and incomes in LAC, going back to the late 1990s.
  - But the studies using income suffer from severe data limitations, given absence of data linking observed parental and adult child incomes that avoid co-residency bias
- Many studies provided TSTSLS estimates
  - Grawe (2004) for Peru (β = 0.67)
  - Ferreira and Veloso (2006) for Brazil (β = 0.58)
  - Dunn (2007) for Brazil (β = 0.69)
  - Nuñez and Miranda (2010) for Chile (β = 0.57)
  - Daza Baez (2021) for Mexico (β = 0.71)
- Two recent studies using administrative data (but thus missing informal sector…)
  - Britto et al. (2022) for Brazil
  - Leites et al. (2022) for Uruguay

2. Inherited inequality: data and methods

- We seek to compensate for these data shortcomings by exploiting available information on other parental and inherited characteristics, i.e., circumstances, available from surveys with retrospective questions.
  - In particular, we use 27 Household surveys covering nine countries, in the 2000 -2015 period
    - From the SEDLAC harmonized database
    - Must contain retrospective questions on parental background, e.g., mother's and father's educational attainment and occupation
- But this I.Op. approach suffers from its own data challenges, chiefly the choice of specification in $y_c = \alpha + C\beta + \varepsilon$ (including interactions) or, equivalently, of the population partition in a non-parametric setting.
  - Trade-off between (downward) omitted-variable bias and (upward) overfitting bias.
  - Use data-driven approaches to select optimal partition (in a well-defined statistical sense)
    - Different algorithms particularly appropriate for ex-ante and ex-post versions of the I.Op. approach.

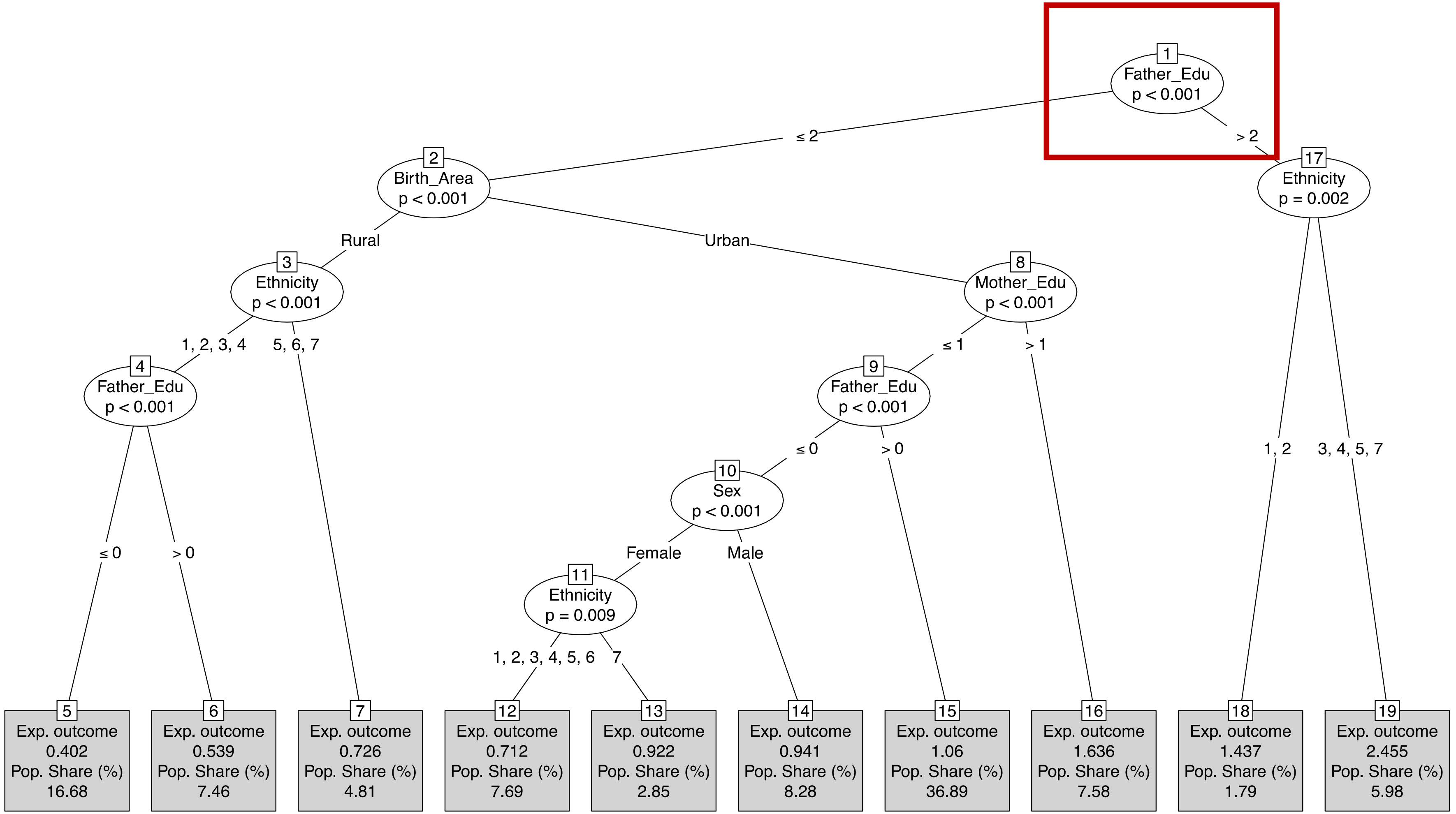
Conditional inference trees and ex-ante inequality of opportunity

Figure 1: Conditional Inference Tree for Bolivia, 2008



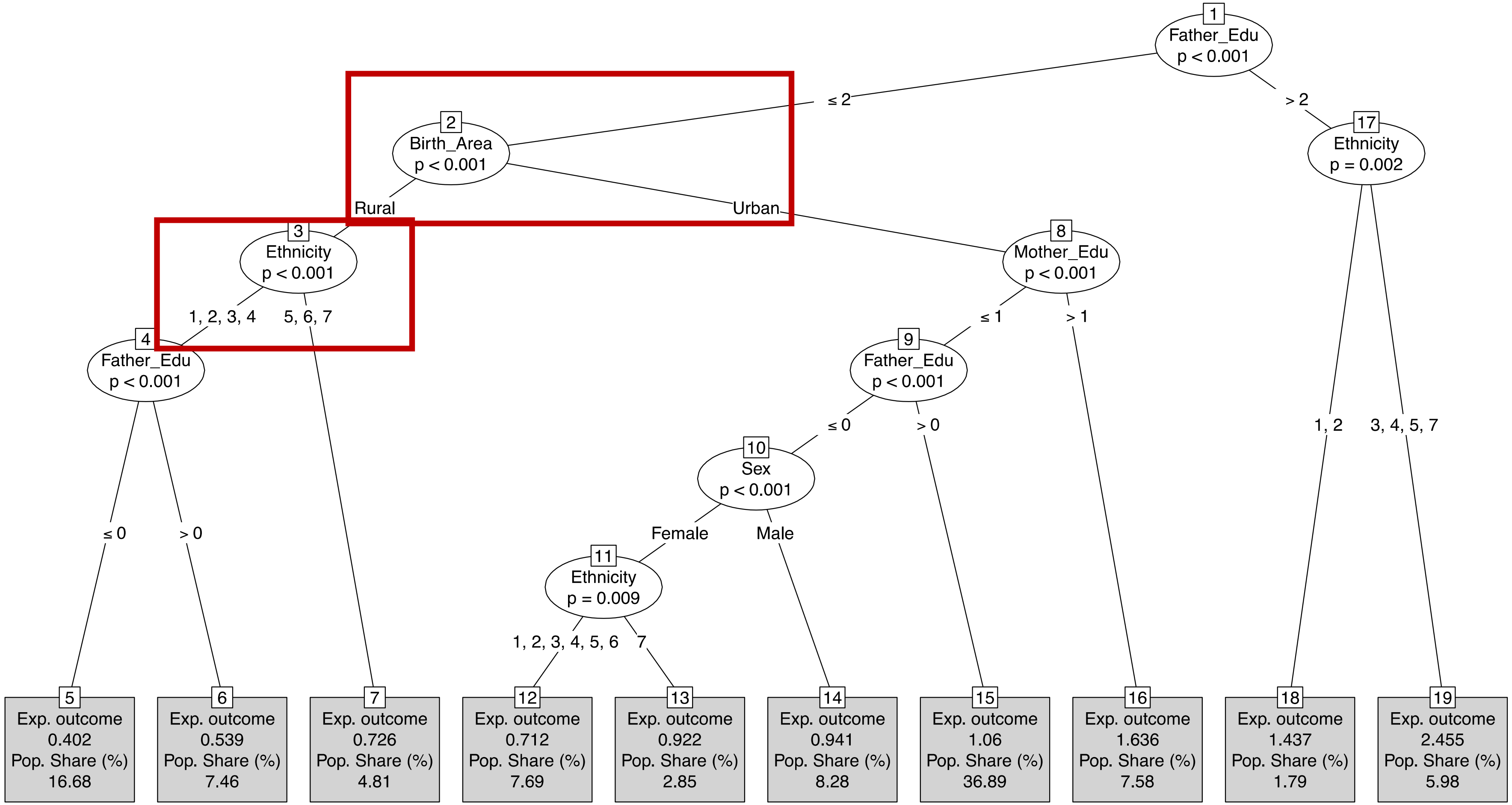
Conditional inference trees and ex-ante inequality of opportunity

Figure 1: Conditional Inference Tree for Bolivia, 2008



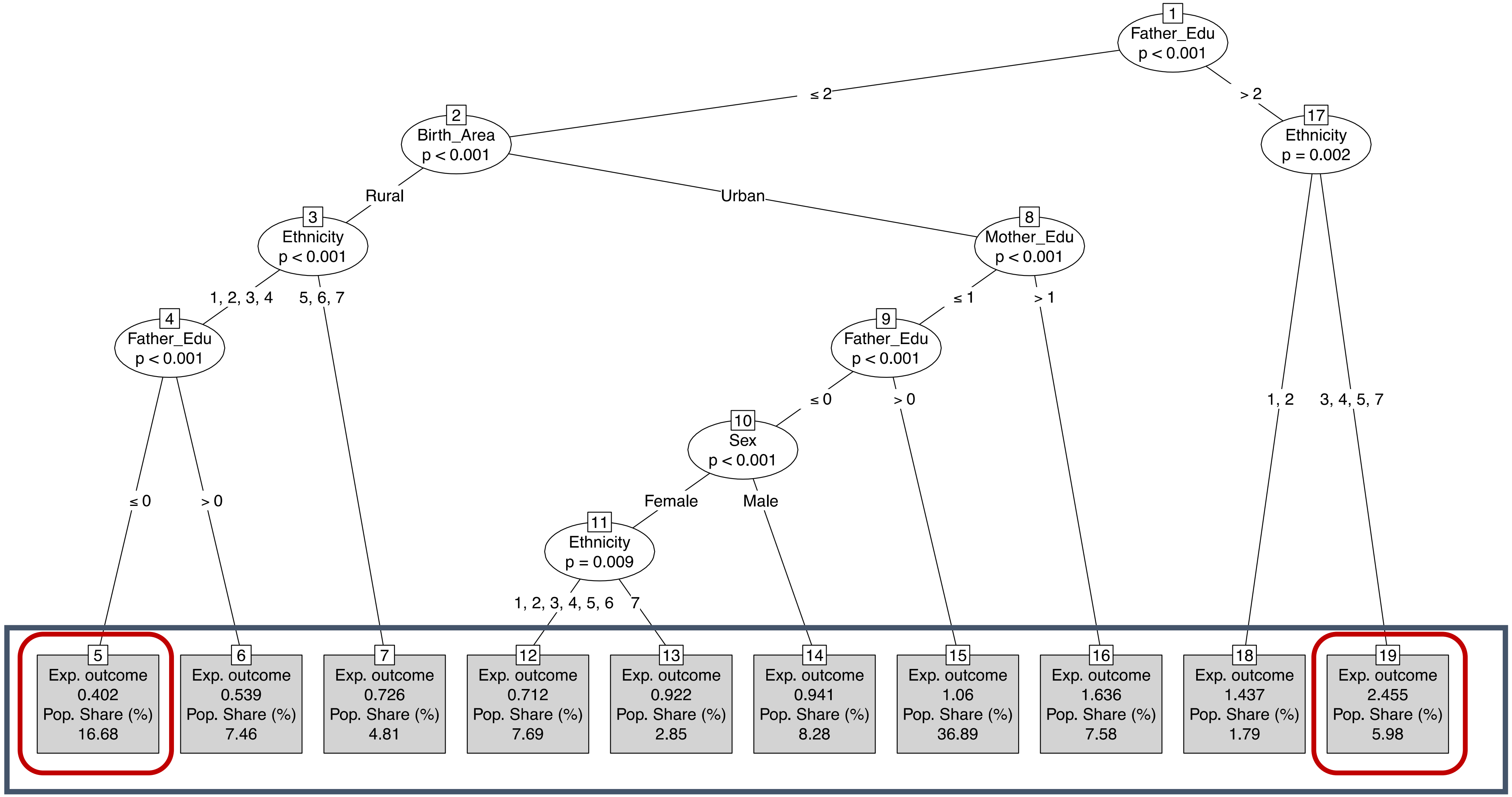
Conditional inference trees and ex-ante inequality of opportunity

Figure 1: Conditional Inference Tree for Bolivia, 2008



Conditional inference trees and ex-ante inequality of opportunity

Figure 1: Conditional Inference Tree for Bolivia, 2008



IOp Gini: 0.22
45% of total

Transformation trees and ex-post inequality of opportunity

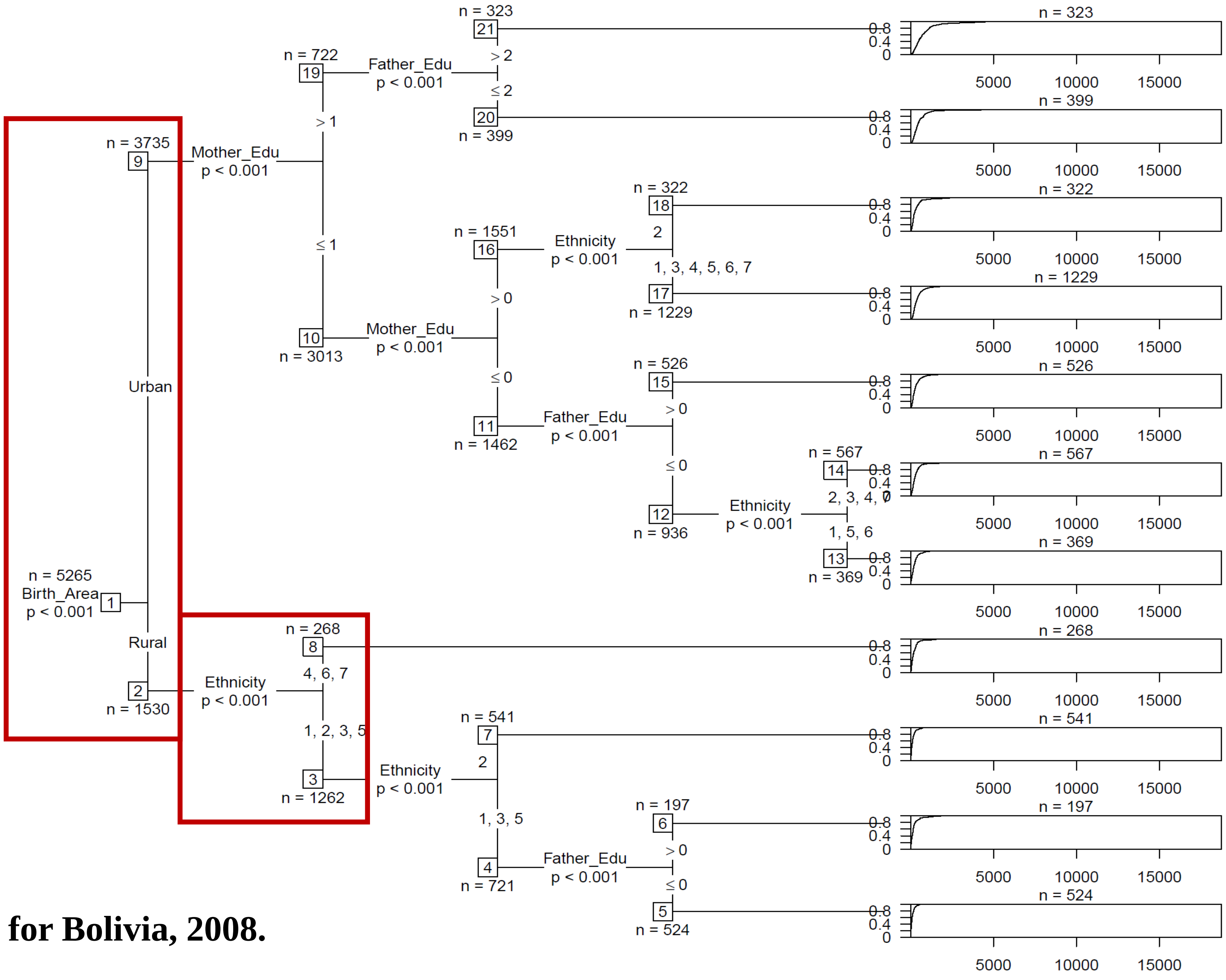
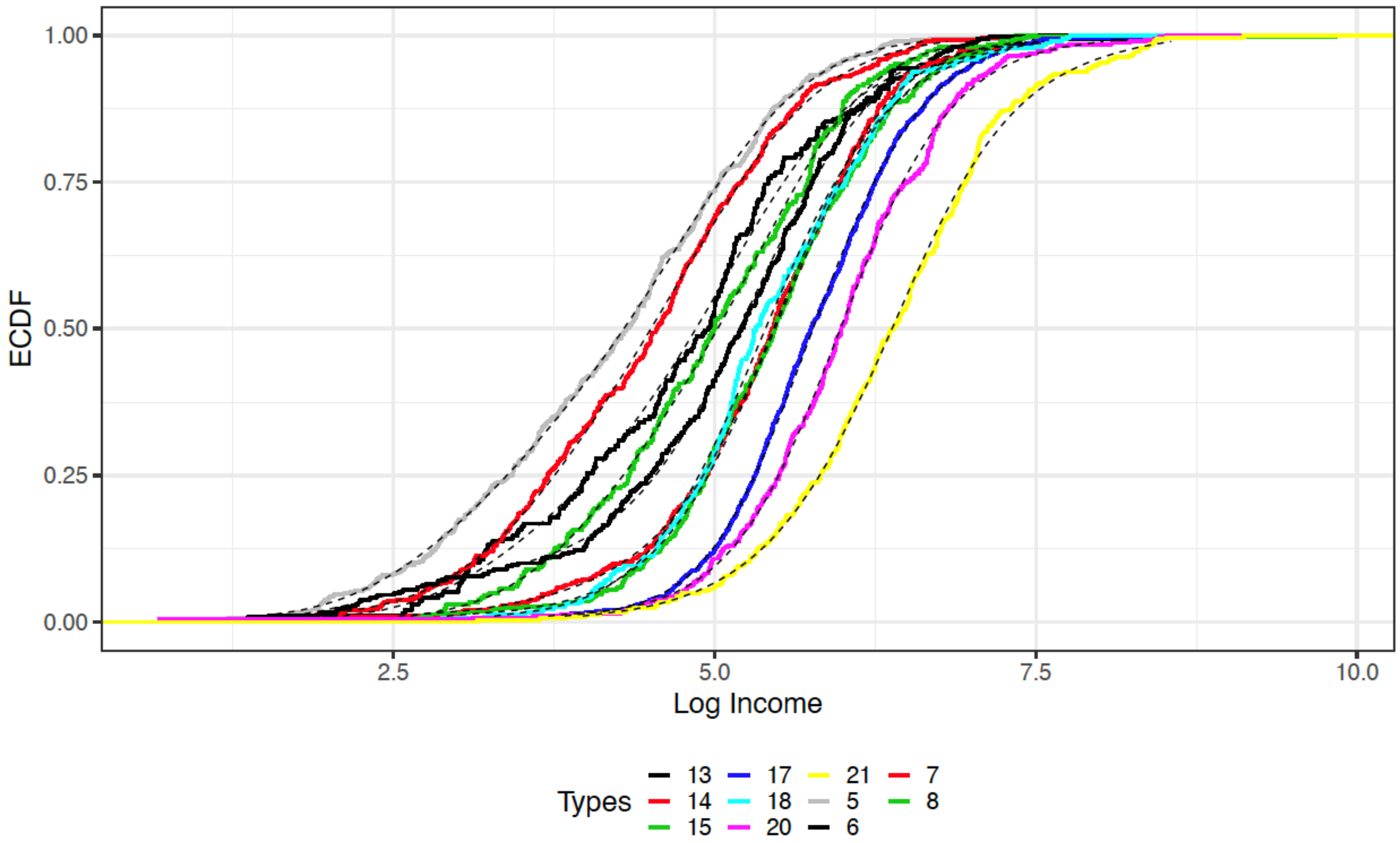


Figure 2: Transformation tree for Bolivia, 2008.

Transformation trees and ex-post inequality of opportunity

Figure 3: Type-specific expected cumulative distribution functions for Bolivia, 2008



Comparisons across methods

Sources:

Ex-ante and ex-post trees computed as above

Education R^2 computed from education data in our surveys

"Approximate" income R^2 computed from TSTSLS estimates in the literature, under the (huge) assumption that marginal standard deviations are equal.

Country/Year	Relative Gini: ex-ante tree	Relative Gini: ex-post tree	Education Rho^2	Approx Income Rho^2
Argentina_2014	0.4377	0.4428	0.239	0.591
Bolivia_2008	0.4450	0.5840		
Brazil_2014	0.5889	0.5779	0.253	0.302
Chile_2006	0.5248	0.4936	0.248	0.325
Chile_2009	0.5492	0.4629	0.283	
Chile_2011	0.5613	0.4935	0.321	
Chile_2015	0.5259	0.4723	0.312	
Colombia_2010	0.5071	0.5221	0.200	
Ecuador_2006	0.4402	0.4774	0.370	1.277
Ecuador_2014	0.5445	0.5766	0.314	
Guatemala_2000	0.4422	0.4732	0.354	
Guatemala_2006	0.5421	0.5383	0.354	
Guatemala_2011	0.6185	0.6290	0.286	
Panama_2003	0.5104	0.4694	0.331	
Panama_2008	0.5521	0.5701	0.337	0.078
Peru_2001	0.5135	0.5721	0.311	0.449
Peru_2005	0.5485	0.5389	0.304	
Peru_2006	0.6038	0.5980	0.298	
Peru_2007	0.5731	0.5678	0.285	
Peru_2008	0.5551	0.5733	0.294	
Peru_2009	0.5337	0.5532	0.284	
Peru_2010	0.5039	0.5299	0.292	
Peru_2011	0.5068	0.5163	0.271	
Peru_2012	0.4937	0.5330	0.265	
Peru_2013	0.5021	0.5397	0.283	
Peru_2014	0.5128	0.5515	0.285	
Peru_2015	0.5122	0.5779	0.273	
Average	**0.524**	**0.535**	**0.294**	**0.504 (0.345)**

Conclusions

- As the study of both inequality and mobility in advanced economies shifts increasingly toward large administrative datasets, the presence of large informal sectors in Latin America hampers our ability to replicate those methods
  - Even after the important and painstaking work of merging and connecting administrative datasets is done, which is currently just beginning in the region.
- Income information from surveys; social security; tax; and national account data are each incomplete in themselves.
  - There is no single, agreed best way to combine them
  - All of which means that there currently is considerable and inevitable uncertainty about levels of income inequality in the region.
  - Trends are more robust: inequality bands
- New machine learning methods can enhance the use of non-income circumstance data for assessing the extent to which inequality is inherited.
  - And even when reliable data on parental income is observable, it is not a sufficient statistic.

Thank you!