Inequality of Opportunity and Intergenerational Persistence in Latin America

Francisco H. G. Ferreira

London School of Economics

(drawing on joint work with François Bourguignon, Paolo Brunori and Guido Neidhöfer)

2022 WIDER-UNIANDES Development Conference

Bogotá, 5 October 2022

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.

LSE The London School of Economics and Political Science

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

IDB Inter-American Development Bank

Inter-American Development Bank

IFS Institute for Fiscal Studies

Institute for Fiscal Studies

Yale University

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

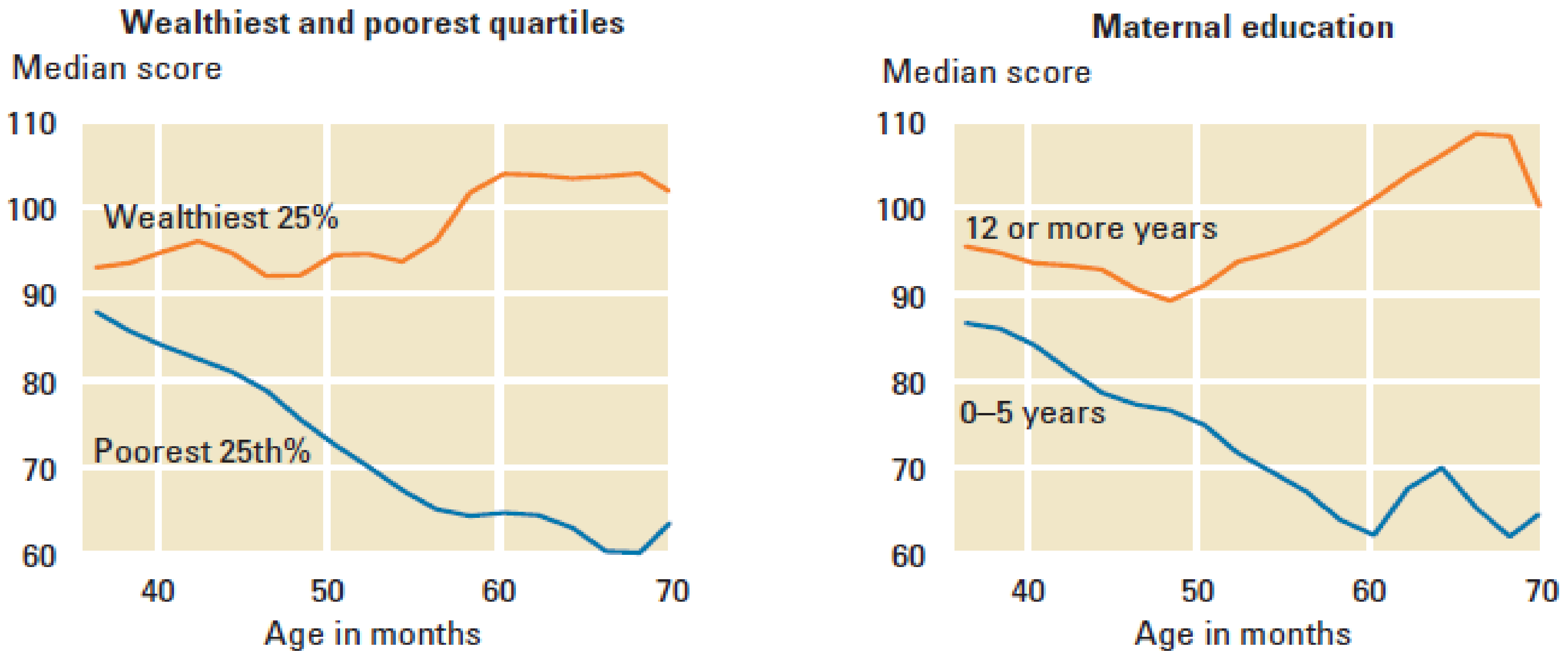
Outline

1. Motivation
2. Review of approaches to measurement
3. A new approach
4. Data
5. Ex-ante inequality of opportunity
6. Ex-post inequality of opportunity
7. Comparisons and conclusions

1. Motivation

- How persistent is socioeconomic advantage across generations?

Figure 2 Opportunities are determined early
Cognitive development for children ages three to five in Ecuador differs markedly across different family backgrounds

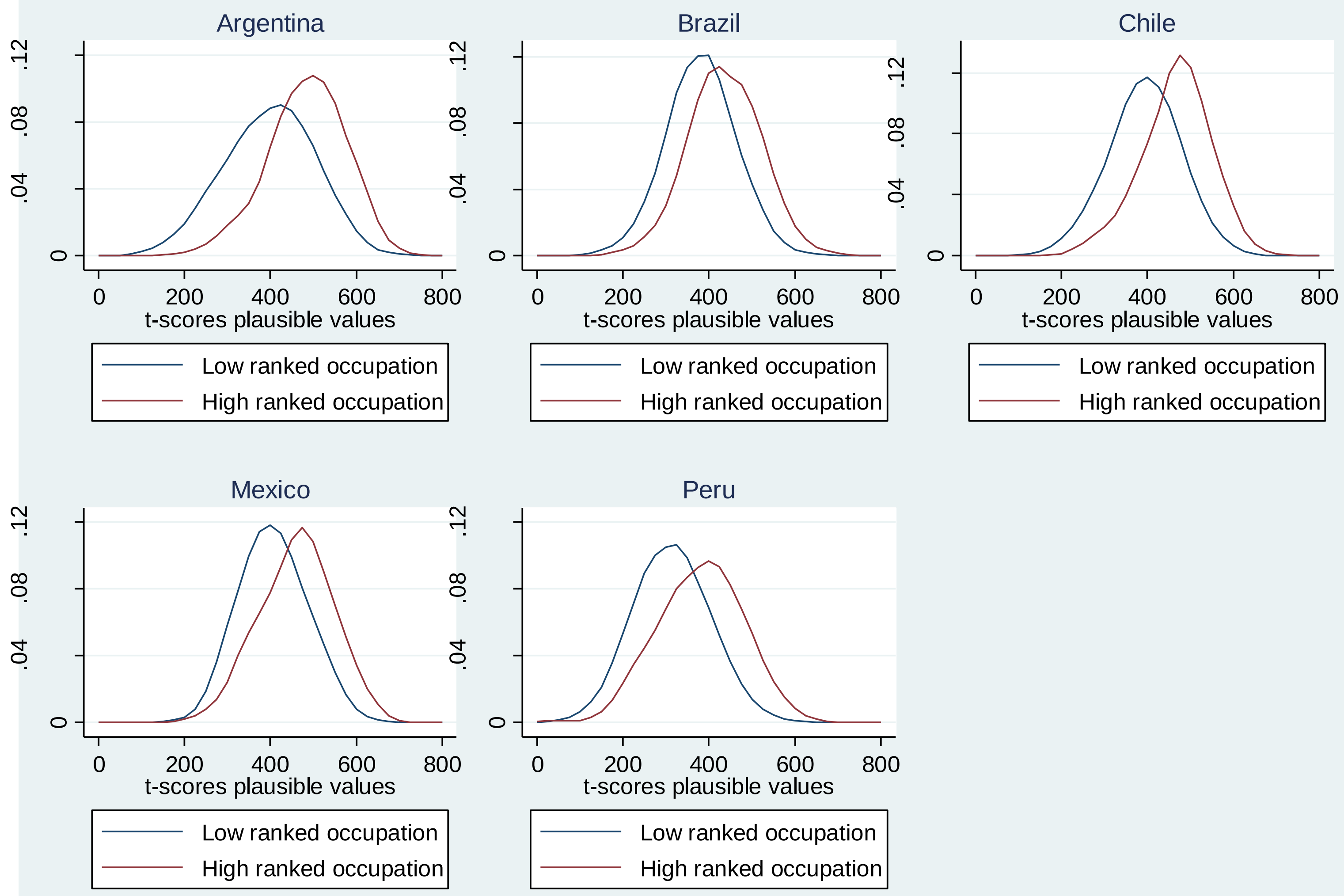


Source: Paxson and Schady (2005a).
Note: Median values of the test of vocabulary recognition (TVIP) score (a measure of vocabulary recognition in Spanish, standardized against an international norm) are plotted against the child's age in months. The medians by exact month of age were smoothed by estimating fan regressions of the median score on age (in months), using a bandwidth of 3.

Source: Paxson and Schady (2007)

1. Motivation

- How persistent is socioeconomic advantage across generations?

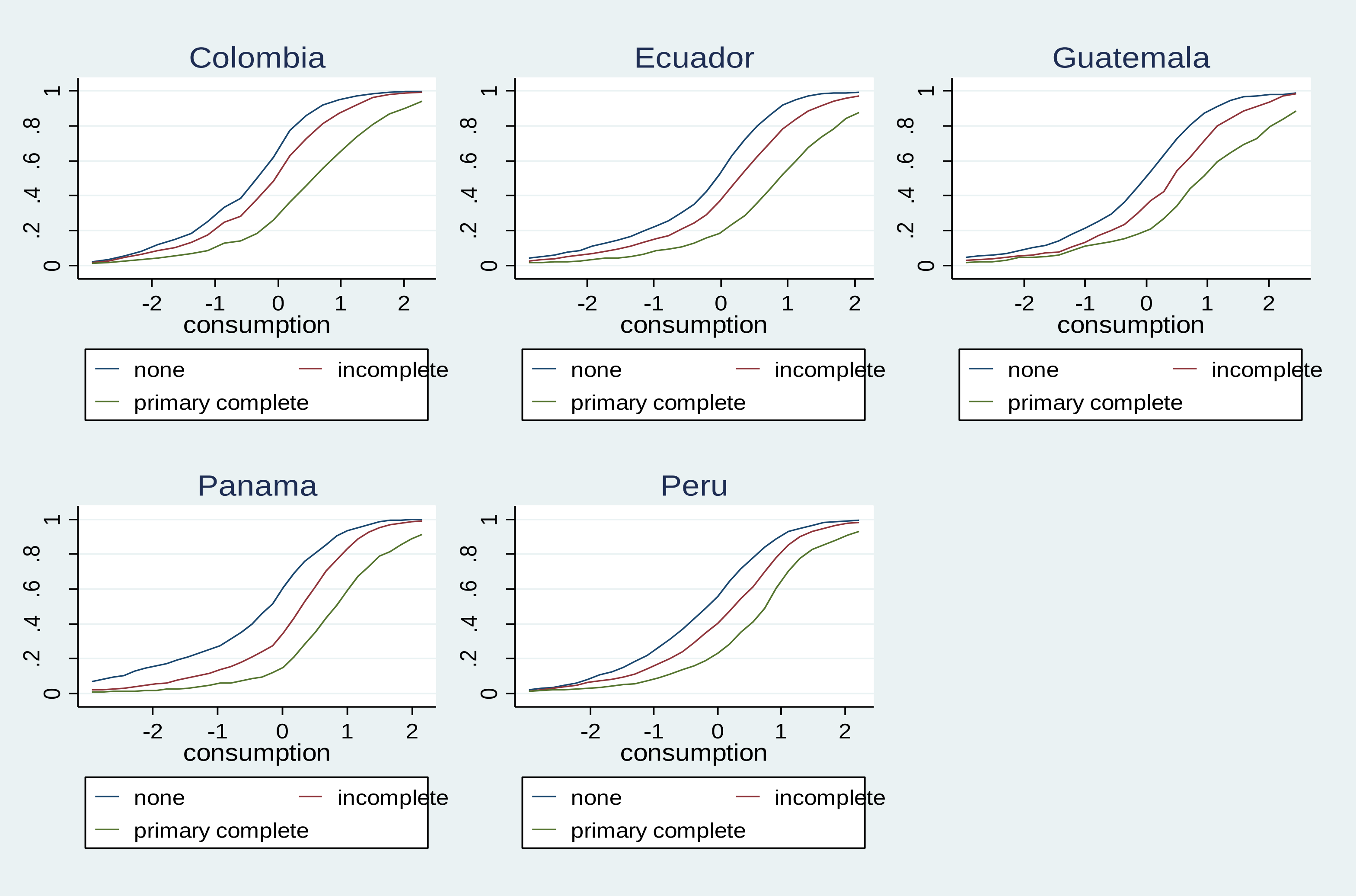


Distributions of reading test scores, conditional on father's occupation (PISA 2006).

Source: Barros, Ferreira, Molinas & Saavedra (2008)

1. Motivation

- How persistent is socioeconomic advantage across generations?



Distributions of per capita household consumption, conditional on mother's education (national household surveys, various years).

Source: Ferreira, and Gignoux (2011)

1. Motivation

- How persistent is socioeconomic advantage across generations?

Juan Vázquez de Coronado y Anaya
Born: 1523 in Salamanca, Spain
Spanish conquistador of Costa Rica

Cuadro V-1

Diputados y Presidentes en la Descendencia de Juan Vázquez de Coronado*

Juan Vázquez de Coronado
Isabel Arias Dávila

Rama A:	2 Presidentes,	48 Diputados
Rama B:		22 Diputados
Rama C:		1 Diputado
Rama D:	3 Presidentes,	26 Diputados
Rama E:		7 Diputados
Rama F:		21 Diputados
Rama G:		4 Diputados
Rama H:		7 Diputados
Rama I:	1 Presidente,	2 Diputados
Rama J:		1 Diputado
Rama K:		1 Diputado
Rama L:		7 Diputados
Rama M:	9 Presidentes,	35 Diputados
Rama N:		1 Diputado
Rama O:	1 Presidente,	10 Diputados
Rama P:	2 Presidentes,	23 Diputados
Rama Q:	4 Presidentes,	12 Diputados
Rama R:	2 Presidentes,	34 Diputados
Rama S:		1 Diputado
Rama T:		2 Diputados
Rama U:	6 Presidentes,	17 Diputados
Rama V:	1 Presidente,	3 Diputados
Total:	31 Presidentes,	285 Diputados

Source: Samuel Z. Stone (1975) – w/ thanks to Fergusson, Robinson and Torres

2. Review of approaches to measurement

- How can the extent of intergenerational persistence be quantified?
  - Two main outcome variables used in the economics literature
    - Income
    - Education
  - Two main approaches – both descriptive:
    - Intergenerational mobility (IGM)
    - Inequality of opportunity (IOp)

2. Review of approaches to measurement

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 seem quite different. Yet, for "origin-independence" mobility and one common measure of IOp:

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

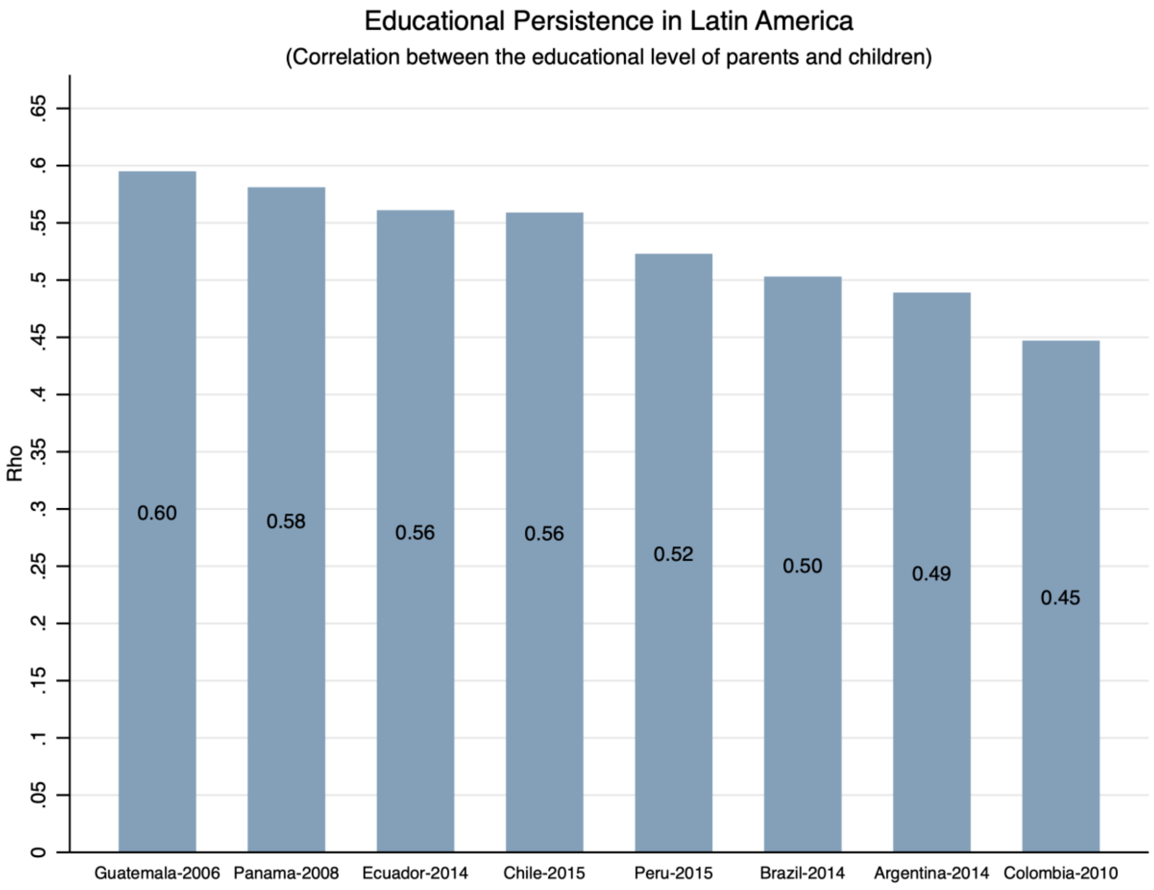
$$y_c = \alpha + C\beta + \varepsilon \qquad IOp^{ea} = \frac{I(\widehat{y_c})}{I(y)}$$

2. Review of approaches to measurement

2.1.a: IGM in education: key studies and findings

- Behrman, Birdsall and Székely (1999)
- Behrman, Gaviria and Székely (2001)
- Hertz et al. (2008)
- Torche (2014)
- Daude and Robano (2015)
- Neidhöfer, Serrano and Gasparini (2018)
- Muñoz (2021)

- Mobility typically lower than in developed countries, particularly for older cohorts (ρ = 0.45 – 0.60 not uncommon)
- Absolute mobility rises for younger cohorts, in part reflecting educational expansions, in countries such as Argentina, Brazil, Costa Rica and Venezuela.
- Relative mobility (e.g., rank correlations) either stable or show very slight improvements.



2. Review of approaches to measurement

2.1.b: IGM in income

- Severe data limitations, given absence of data linking parental and adult child incomes that avoid co-residency bias
- Many studies provided TSTSLS estimates
  - Grawe (2004) for Peru ($\beta = 0.67$)
  - Ferreira and Veloso (2006) for Brazil ($\beta = 0.58$)
  - Dunn (2007) for Brazil ($\beta = 0.69$)
  - Nuñez and Miranda (2010) for Chile ($\beta = 0.57$)
  - Daza Baez (2021) for Mexico ($\beta = 0.71$)
- One recent study using administrative data (thus missing informal sector…)
  - Leites et al. (2022) for Uruguay
- Some recent work exploring three generations
  - Celhay and Gallegos (2015, 2022)

2. Review of approaches to measurement

2.2.a: IOp in income

- Incorporates other circumstance variables, such as parental occupation and race
- Some substantial shares reported, typically interpreted as lower-bound.
  - Bourguignon et al. (2007)
  - Barros et al. (2009)
  - Ferreira and Gignoux (2011)

2.2.b: IOp in education (test scores)

- Gamboa and Waltenberg (2012)
- Ferreira and Gignoux (2014)

TABLE 6

SCALAR INDICES OF INEQUALITY OF OPPORTUNITY

	Brazil	Colombia	Ecuador	Guatemala	Panama	Peru
Panel A: Household income (per capita)						
Total inequality (E_0)	0.692	0.572	0.580	0.593	0.630	0.557
	(0.013)	(0.033)	(0.028)	(0.036)	(0.029)	(0.022)
Non-parametric estimates						
IOL	0.227	0.144	0.164	0.213	0.213	0.163
	(0.008)	(0.023)	(0.022)	(0.031)	(0.024)	(0.015)
IOR	0.329	0.252	0.283	0.359	0.338	0.293
	(0.008)	(0.026)	(0.023)	(0.030)	(0.026)	(0.018)
Parametric estimates						
IOL	0.223	0.133	0.150	0.199	0.190	0.156
	(0.008)	(0.019)	(0.020)	(0.028)	(0.023)	(0.014)
IOR	0.322	0.232	0.259	0.335	0.301	0.279
	(0.008)	(0.023)	(0.023)	(0.030)	(0.028)	(0.018)
Panel B: Household consumption expenditures (per capita)						
Total inequality (E_0)		0.462	0.359	0.415	0.381	0.351
		(0.018)	(0.015)	(0.025)	(0.018)	(0.013)
Non-parametric estimates						
IOL		0.123	0.124	0.221	0.156	0.123
		(0.015)	(0.013)	(0.024)	(0.016)	(0.010)
IOR		0.265	0.346	0.532	0.409	0.351
		(0.021)	(0.021)	(0.023)	(0.025)	(0.018)
Parametric estimates						
IOL		0.114	0.117	0.213	0.144	0.119
		(0.014)	(0.012)	(0.022)	(0.015)	(0.009)
IOR		0.247	0.326	0.514	0.377	0.339
		(0.021)	(0.022)	(0.022)	(0.026)	(0.017)

Notes: Sample: household heads and spouses, aged 30–49, with positive income and information on a set of circumstances; bootstrap standard errors (taking into account stratification and clustering) in parentheses; father's occupation missing for Colombia and Peru.

Source: Ferreira and Gignoux (2011)

2. Review of approaches to measurement

These early I.Op. estimates were based on plausible, but arbitrary, partitions of the population

		Brazil	Colombia	Ecuador	Guatemala	Panama	Peru
Ethnicity							
	category 1	self reported white ethnicity	Other	self-reported ethnicity: white, mixed blood (“mestizo”) or other	European maternal language		European maternal language
	category 2	self reported black (“negro”) and mixed blood (“pardo”) ethnicity	self-reported minority ethnicity: “indígena, gitano, archipiélago o palenquero”	self-reported ethnicity: indigenous, black (“negro” or “mulato”).	indigenous maternal language	speaks indigenous language	indigenous maternal language
Father's occupation							
	category 1	agricultural worker	Missing	agricultural worker or domestic worker	agricultural worker	agricultural worker	missing
	category 2	Other		Other	other	other	
Mother’s and father’s education							
	category 1	None or unknown	none or unknown	none or unknown	none or unknown	none or unknown	none or unknown
	category 2	completed grade 1 to 4	primary incomplete	Primary	primary incomplete	primary	primary incomplete
	category 3	completed grade 5 or more	primary complete or more	secondary or more	primary complete or more	secondary or more	primary complete or more
Birth region							
	category 1	Sao Paulo & Federal district	departments at the periphery	Sierra & Amazonia provinces	Guatemala city, North-East departments and El Petén	cities and intermediate urban centers	Inland non-southern departments
	category 2	South East, Center-West & South	Central departments(a)	Costa & Insular provinces	North & North-West departments	other urban centers	Southern and other costal departments
	category 3	North-East, North or missing	Bogota, San Andres and Providencia islands and foreign country	Pichincha province (with Quito) & Azuay province	South-East, South-West & Center departments	rural areas	Arequipa, Callao & Lima

Source: Ferreira and Gignoux (2011)

3. A new approach (to IOp for income)

- Two alternative conceptual views of inequality of opportunity:
  - Ex-ante:
    - Equality of opportunity attained when $\mu^k(y) = \mu^l(y), \forall l, k | T_k \in \Pi, T_l \in \Pi$
    - Inequality of opportunity can be measured as the between-group component of the GE-decomposition by population subgroups.
  - Ex-post:
    - Equality of opportunity attained when $F^k(y) = F^l(y), \forall l, k | T_k \in \Pi, T_l \in \Pi$
    - Inequality of opportunity measured as some suitable aggregation $IOp = \int_{q=0}^{1} w_q I_q(y_{qc})$ where $y_{qc} = F^{-1}(q|C = c)$
      - Need estimates of the type-specific quantile functions

- In both approaches, a population partition (into types) is a key first step.

3. A new approach

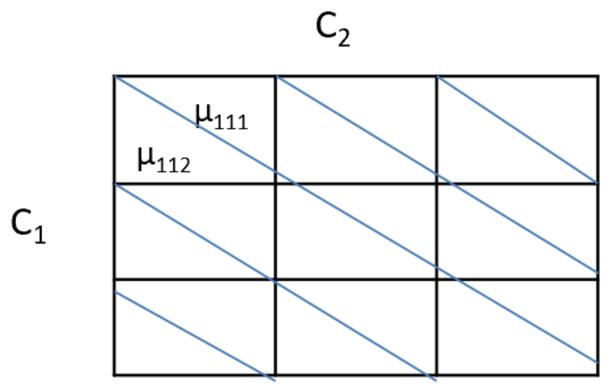
- But how should the population be partitioned?
- Consider Bolivia (2008) in our data set:
  - N = 6,071 observations
  - Circumstances:
    - Sex (2 categories)
    - Ethnicity (7 categories)
    - Occupation of father and mother (9 categories each)
    - Education of father and mother (4 categories each)
  - Number of potential types: 2 x 7 x 9 x 9 x 4 x 4 = 18,144

3. A new approach

- When I.Op. is measured using a sample drawn from the population (as is usually the case), two competing biases may be at play:

1. Downward bias from omitted (unobserved) circumstances
    - Ferreira and Gignoux (2011)

	C_2		
	μ_{11}	μ_{12}	μ_{13}
C_1	μ_{21}	μ_{22}	μ_{23}
	μ_{31}	μ_{32}	μ_{33}



2. Upward bias from overfitting
    - Sampling variation around sub-group parameter estimates explodes as cell sizes become too small. (Brunori, Peragine and Serlenga, 2018)

3. A new approach

- Following Hothorn et al. (2006) and Hothorn and Zeileis (2021), we use adaptive local maximum likelihood methods to:

    1. Select the optimal partition of the population
    2. Estimate features of the conditional distribution within groups
        - For the ex-ante approach, focus on differences in means between types
        - For the ex-post approach, consider differences between quantile functions

- Spirit: given a data set, use as flexible a statistical approach as possible to model its distributional structure

3. A new approach: ex-ante

- Follow Brunori, Hufe and Mahler (2021) in using conditional inference trees and random forests (Hothorn et al., 2006) to select partitions:

1. Given a set of circumstance variables and categories, the algorithm tests the correlation between the outcome and each circumstance. If the Bonferroni-adjusted p-value of the correlation test is higher than the chosen critical value α, one exits the algorithm.
2. If the null hypothesis is rejected, the variable with the smallest Bonferroni-adjusted p-value is selected as the first splitting variable $[c]$.
3. The algorithm then considers how circumstance $[c]$ can be used to partition the sample into two subsamples [C]. Among all possible binary partitions, it computes the p-value for the null hypothesis that the statistic of interest (e.g., the mean) in the two sub-samples is identical.
4. [C]* is chosen as $[C]^* = \{[C]: argmin\ p^{[C]}$ That is to say: when there are n > 2 categories for a particular circumstance variable, the categories are divided into the two groups that are least likely to have the same (say) mean.
5. Repeat steps 1 – 4 for each node (sub-sample), until one has exited everywhere

3. A new approach: ex-post

- Follow Brunori, Ferreira and Salas-Rojo (2022) in using transformation trees (Hothorn and Zeileis, 2021) to select partitions and estimate type-specific quantile functions:
- Key assumption (for the ex-post case): there exists a sufficiently good parametric approximation to $F(y_{qc}|C = c)$. In the limit:

$$F(y_{qc}|C = c) = F\left(y_{qc}, \theta(c)\right), \theta: \mathbb{C} \rightarrow \Theta$$

- If this holds, then the problem is to select:

$$\hat{\theta}^N(c) = \arg\max_{\theta \epsilon \Theta} \sum_{i=1}^{N} w_i(c)\ell_i(\theta)$$

$$w_i(c) = \sum_{b=1}^{B} I(c \in \mathcal{B}_b \wedge c_i \in \mathcal{B}_b)$$

And, using Bernstein polynomials to estimate the conditional distributions within groups yields the following local log-likelihood function: $\ell_i(\theta) = \log\left[f_z\left(a(y)^T\theta\right)\right] + \log\left(a(y)^T\theta\right)$

3. A new approach: ex-post

- In practice:

  1. set a critical value α and a polynomial order P

  2. estimate the unconditional distribution with a polynomial approximation

  3. test the null hypothesis of polynomial parameter stability for all possible partitions based on C and store $p-values$.

  4. if all Bonferroni-adjusted $p-value > \alpha$, stop the algorithm

  5. otherwise, choose the variable and the splitting value producing the smallest $p-value$ to obtain two subgroups. Estimate the conditional distributions in each with a polynomial approximation.

  6. repeat step 3-5 for the resulting subgroups

- For both CI and transformation trees, **random forests** (or equivalent) can help reduce the variance of the tree estimators

4. Data

28 Household surveys covering nine countries

From the SEDLAC harmonized database

1994 -2017

Must contain retrospective questions on parental background, e.g., mother's and father's educational attainment and occupation

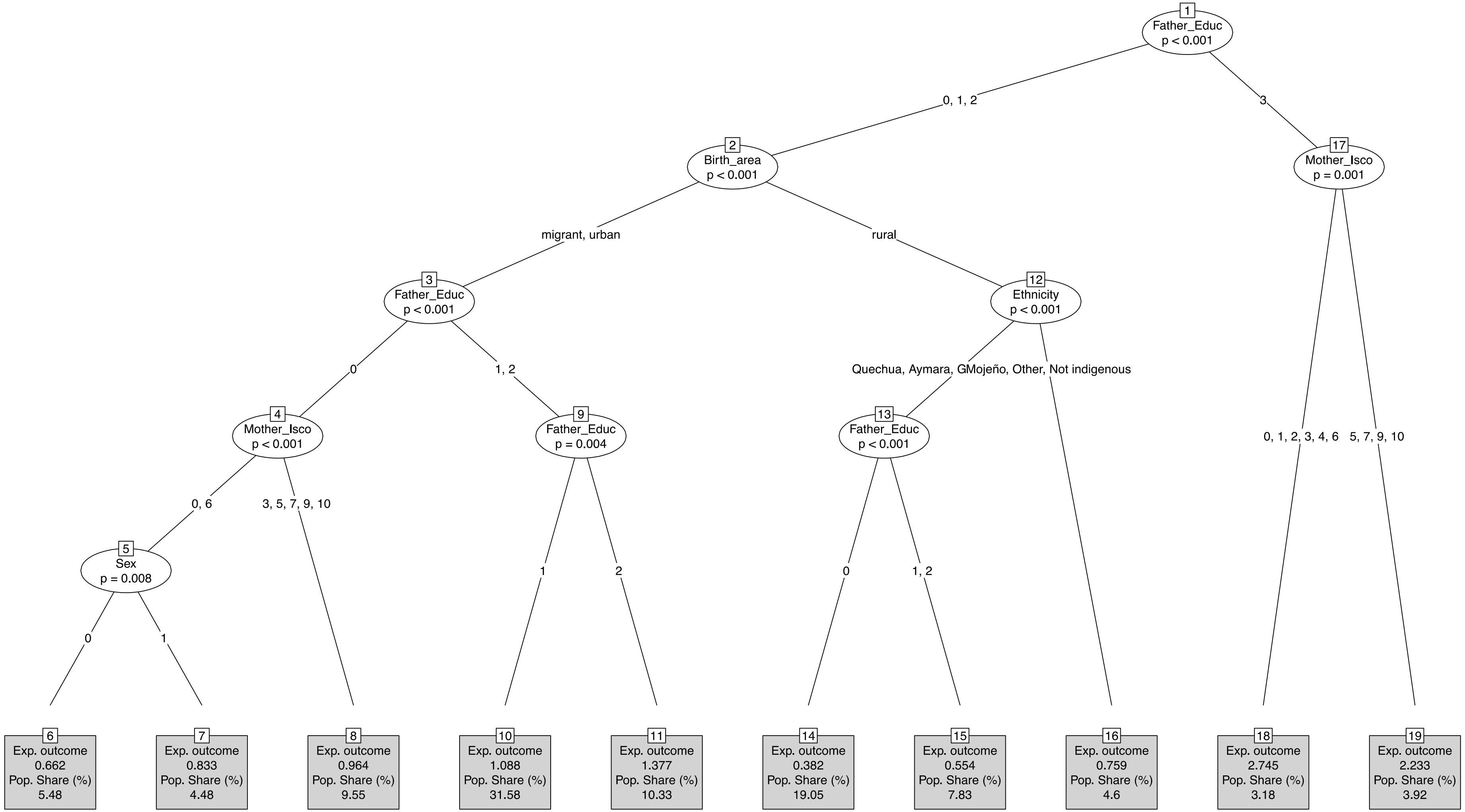
Age range restricted to "central 80% of working age"

Country	Survey Wave	Circumstances	Parents' information asked of :	Original sample	New Age Range **	Final Sample Size ^	Relative sample size (K / H) ^^	Significant difference in mean income*
Argentina	2014	Sex, race or ethnicity, place of birth, father's and mother's education, father's occupation	Household head and partner	13,358	29 - 71	5,481	41.0%	*
Bolivia	2008	Sex, race or ethnicity, father's and mother's education, father's occupation	All individuals aged 12 to 65 years old	10,149	15 - 54	6,071	59.8%	
Brazil	2014	Sex, race or ethnicity, place of birth, father's and mother's education, father's and mother's occupation	One randomly chosen individual older than 15 years old per household.	60,629	22 - 69	24,873	41.0%	*
Chile	2006	Sex, race or ethnicity, place of birth, father's and mother's education, father's and mother's occupation (only 2009)	Household head and partner	123,905	32 - 70	66,231	53.5%	*
	2009			118,069	32 -72	51,088	43.3%	*
	2011			95,694	31 - 72	45,824	47.9%	*
	2013			107,006	31 - 73	60,350	56.4%	*
	2015			133,597	31 -73	76,838	57.5%	*
Colombia	2010	Sex, race or ethnicity (except: 2003, 2008 and 2011), father's and mother's education	All individuals in the household	50,071	15 -61	31,185	62.3%	*
Ecuador	2006	Sex, race or ethnicity, place of birth, father's and mother's education, father's and mother's occupation	All individuals in the household	41,251	15 - 62	24,623	59.7%	*
	2014			83,508	15 - 64	49,896	59.7%	*
Guatemala	2000	Sex, race or ethnicity, place of birth, father's and mother's education, father's and mother's occupation (only 2000)	All individuals older than 12 years old	23,058	16 - 59	13,070	56.7%	
	2006			43,236	16 - 60	27,614	63.9%	
	2011			44,040	16 - 60	27,950	63.5%	
Panama	2003	Sex, race or ethnicity (except 2008), place of birth (except 2003), father's and mother's education, father's and mother's occupation (except 2003)	All individuals in the household	17,374	19 - 65	12,189	70.2%	
	2008			18,496	19 - 65	9,688	52.4%	*
Peru	2001	Sex, race or ethnicity (except: 2002, 2003, 2004 and 2005), place of birth, father's and mother's education	Household head and partner	28,112	28 - 67	19,470	69.3%	
	2005		Household head	19,895	31 - 72	12,354	62.1%	
	2006			20,577	32 -72	11,785	57.3%	
	2007			22,204	31 - 72	13,419	60.4%	
	2008			21,502	31 - 72	12,887	59.9%	
	2009			21,753	32 -73	12,989	59.7%	
	2010			21,496	32 -73	12,813	59.6%	
	2011			24,809	33 - 74	14,643	59.0%	
	2012			25,091	34 - 74	14,834	59.1%	
	2013			30,453	34 - 74	17,717	58.2%	
	2014			30,848	34 - 74	17,780	57.6%	
	2015			32,188	33 - 74	18,473	57.4%	

* A statistically significant difference between the mean of the equivalized household income in the complete sample and the final sample. Significance level 5%.

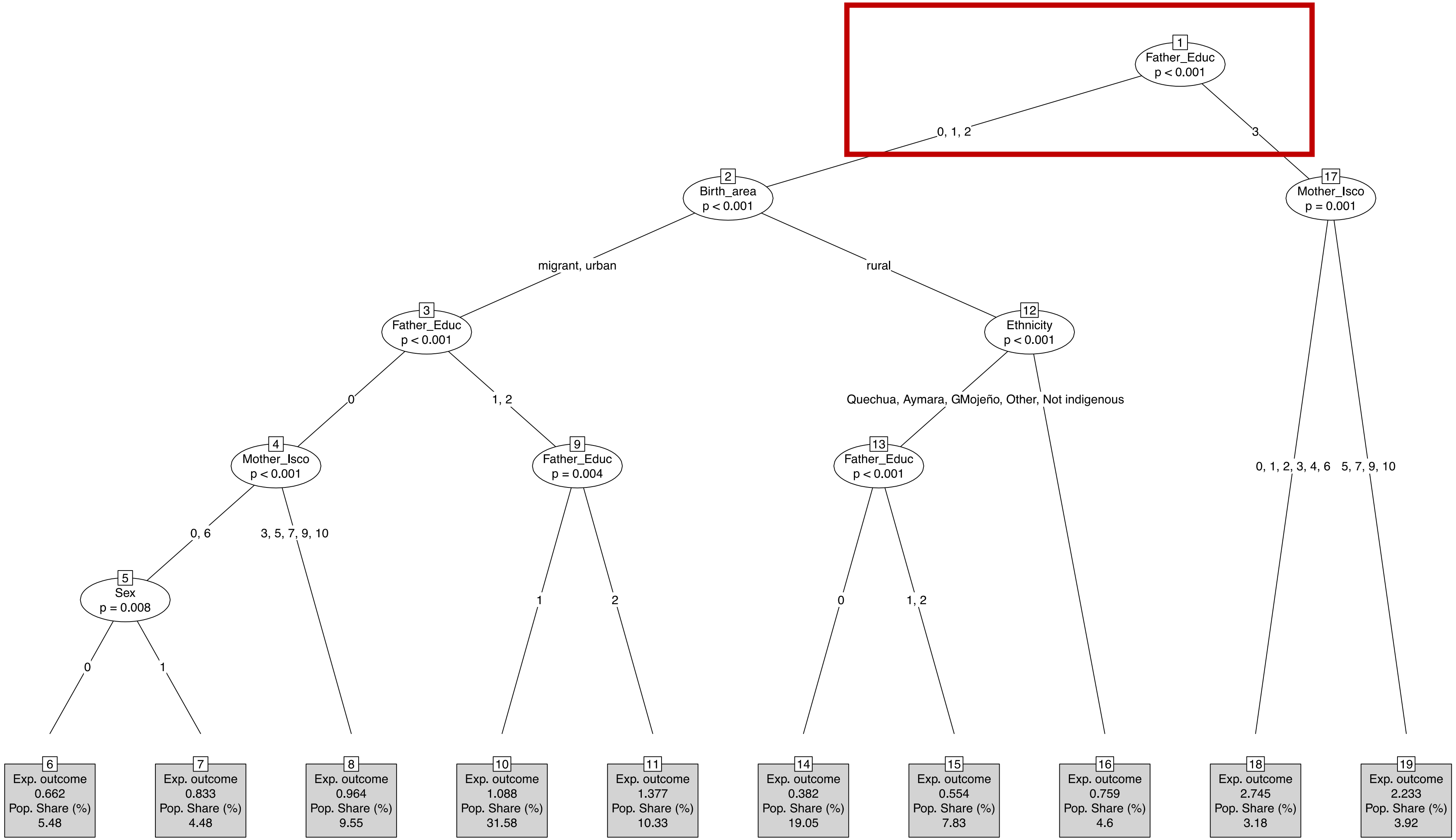
5. Ex-ante Inequality of Opportunity

Figure 1: Conditional Inference Tree for Bolivia, 2008



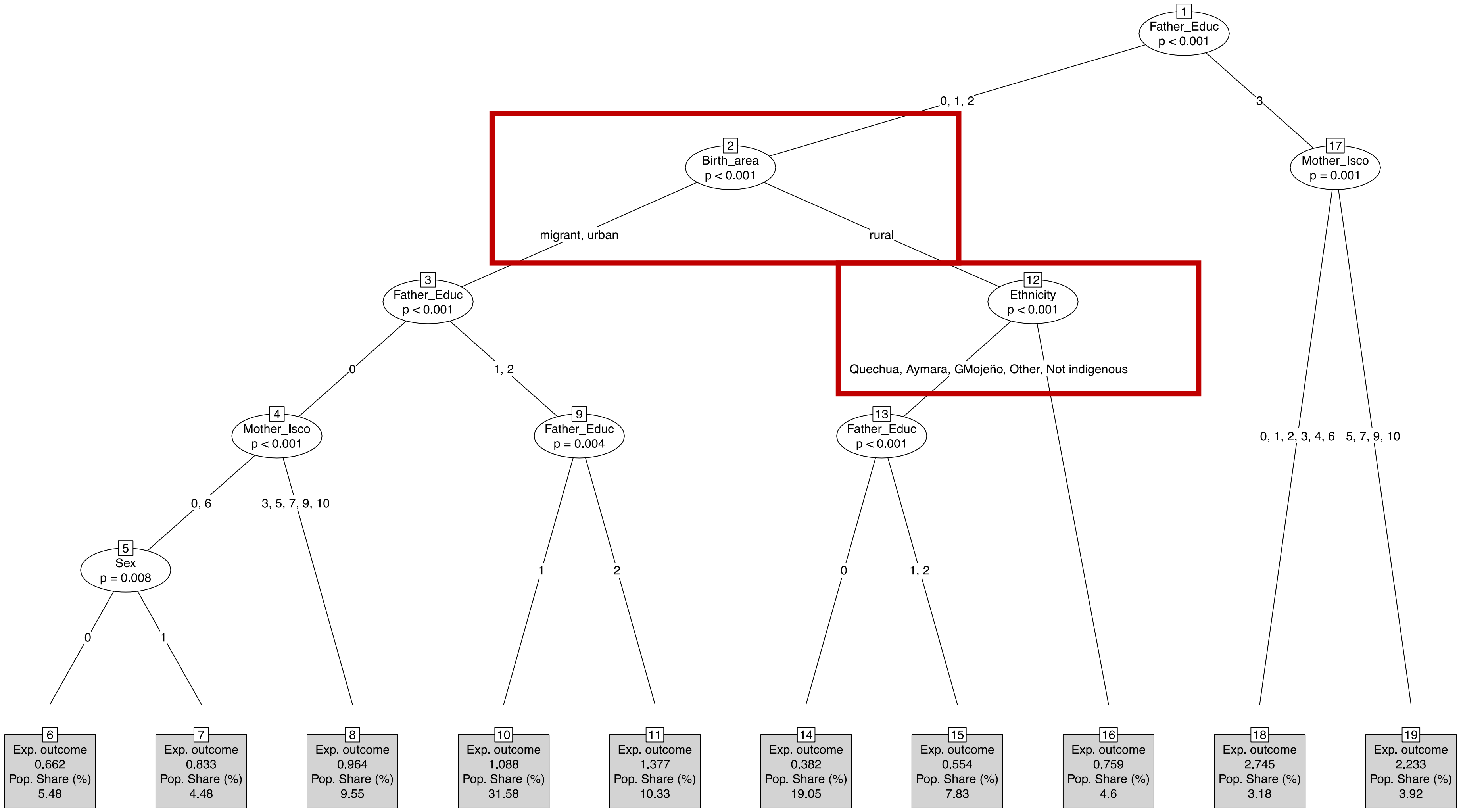
5. Ex-ante Inequality of Opportunity

Figure 1: Conditional Inference Tree for Bolivia, 2008



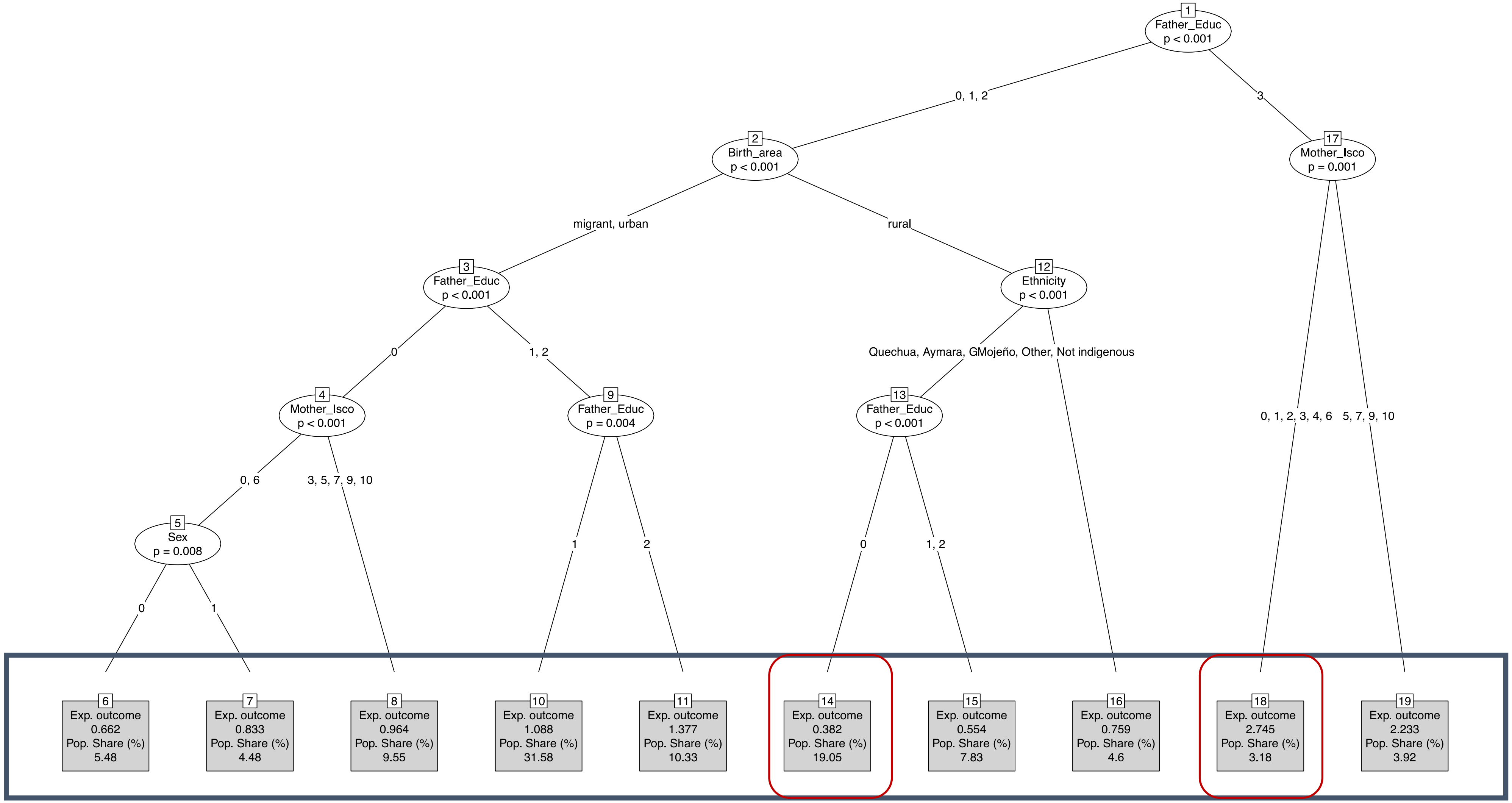
5. Ex-ante Inequality of Opportunity

Figure 1: Conditional Inference Tree for Bolivia, 2008



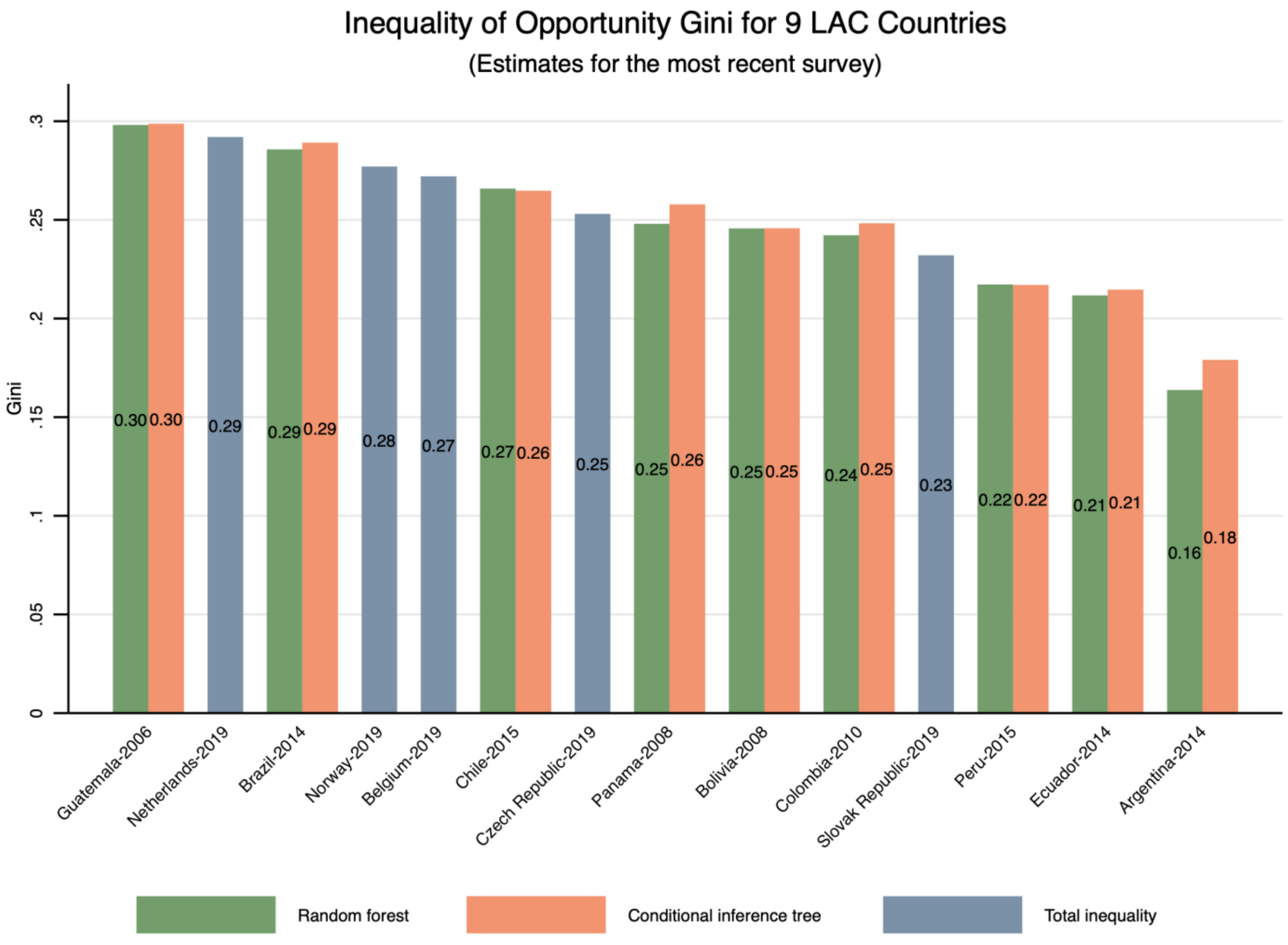
5. Ex-ante Inequality of Opportunity

Figure 1: Conditional Inference Tree for Bolivia, 2008



5. Ex-ante Inequality of Opportunity

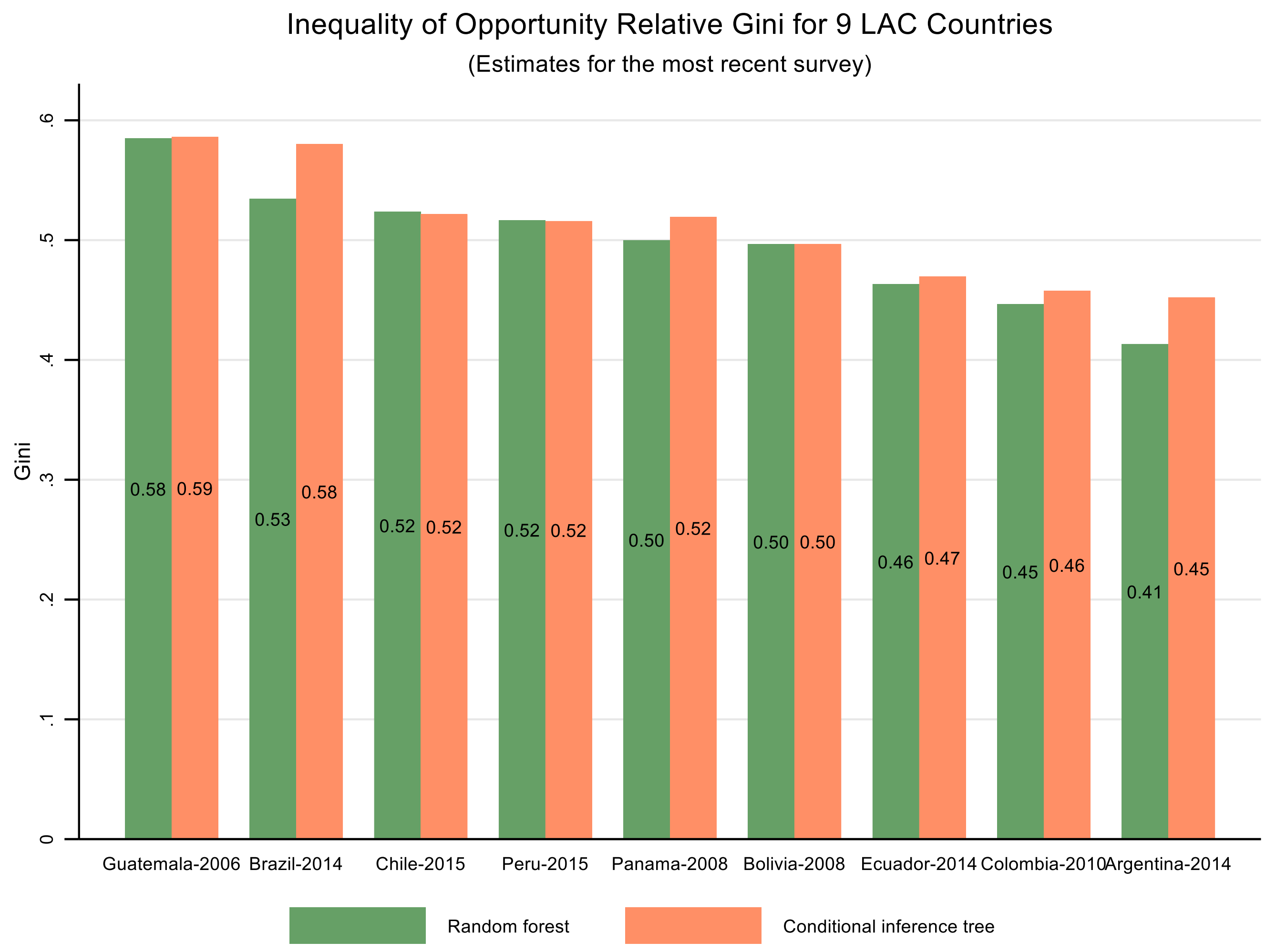
Despite parsimonious partitions, IOp levels in LAC are higher than total inequality in some countries



Source: World Development Indicators online (21 August 2022) for Belgium, Czech Republic, Netherlands, Norway, Slovak Republic.

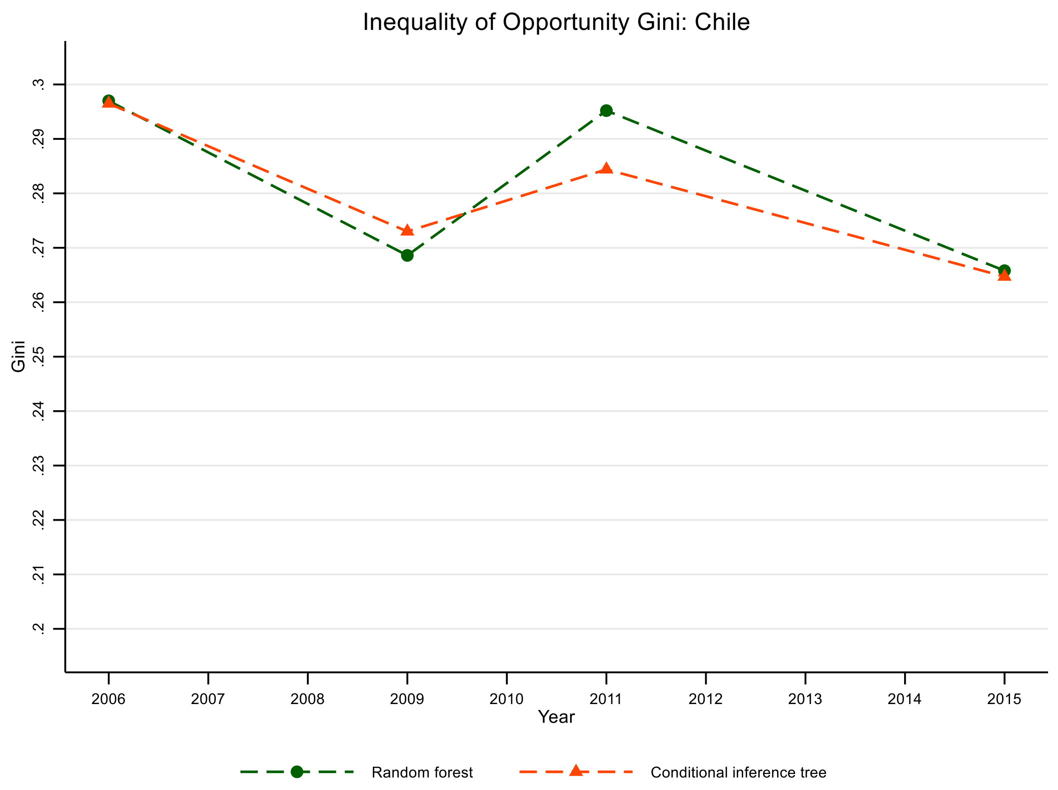
5. Ex-ante Inequality of Opportunity

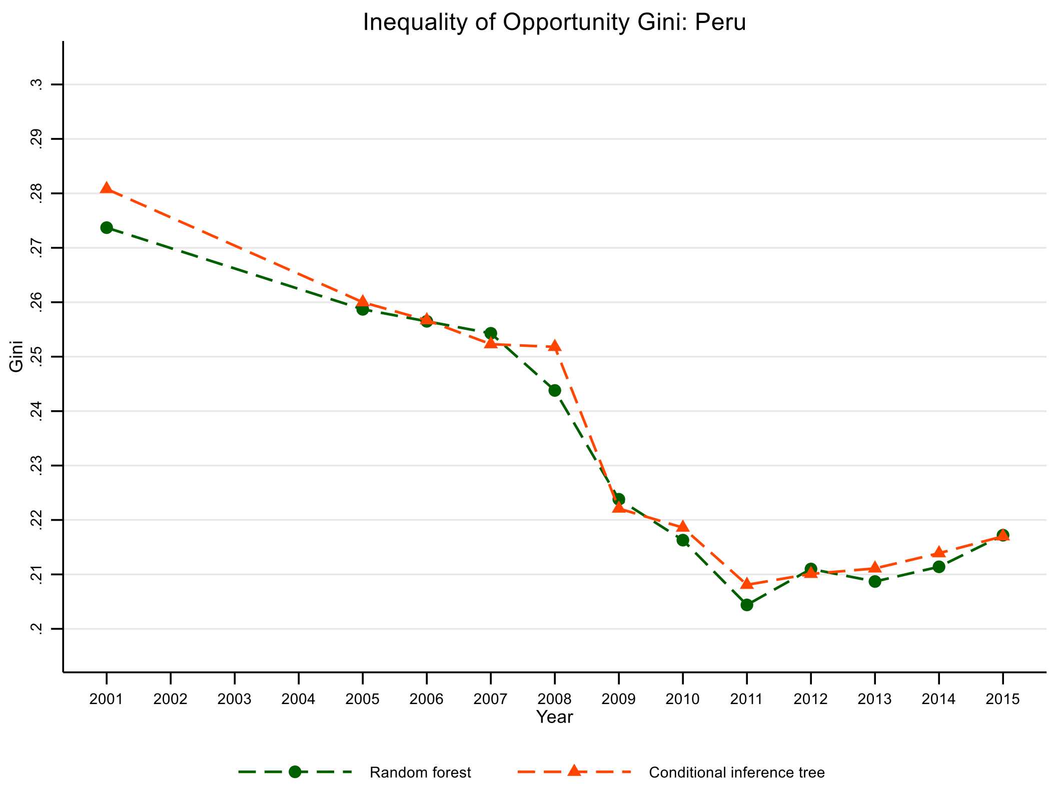
Despite parsimonious partitions, IOp typically accounts for over half of total HEY inequality (for the Gini)



5. Ex-ante Inequality of Opportunity

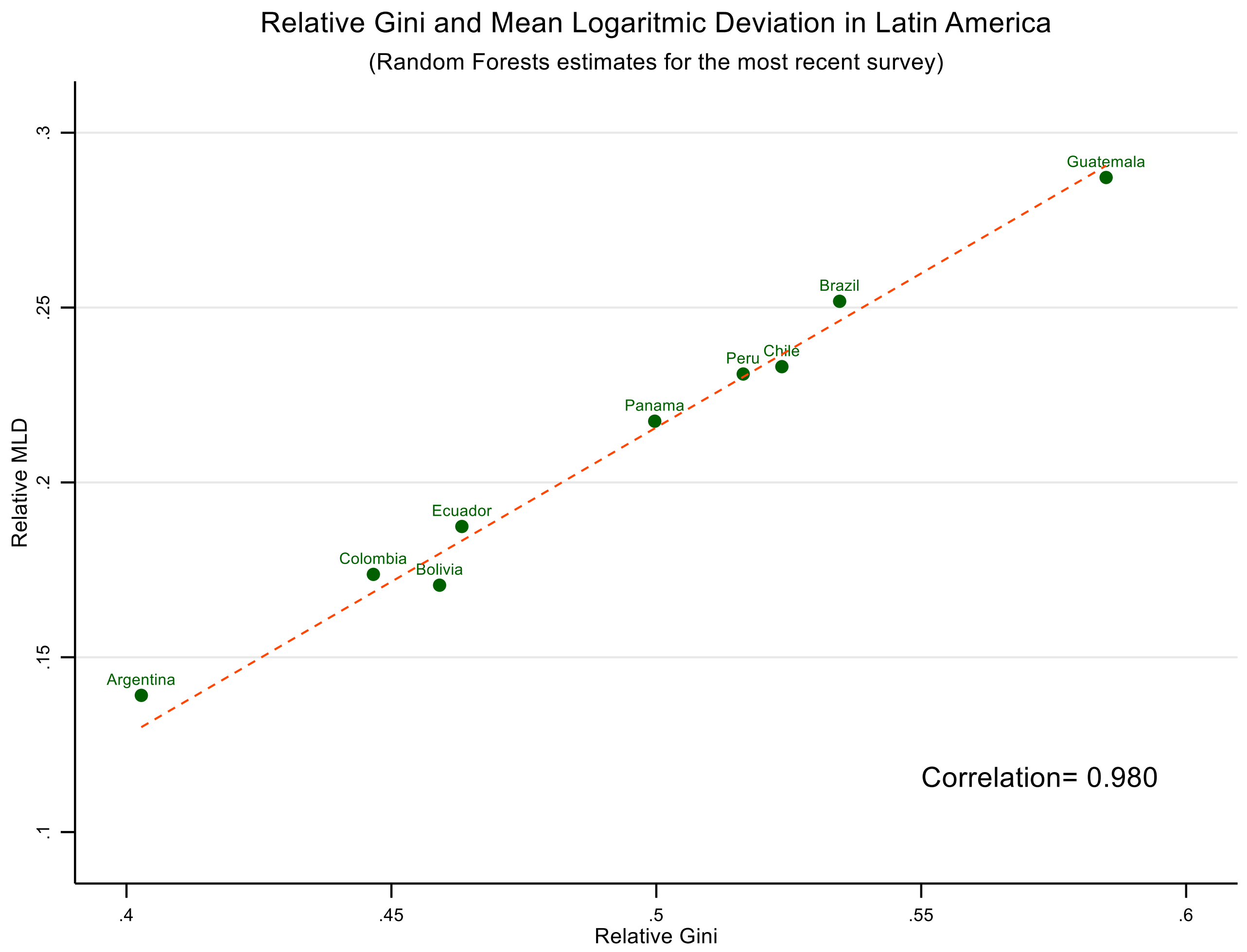
Some time series for Chile and Peru





5. Ex-ante Inequality of Opportunity

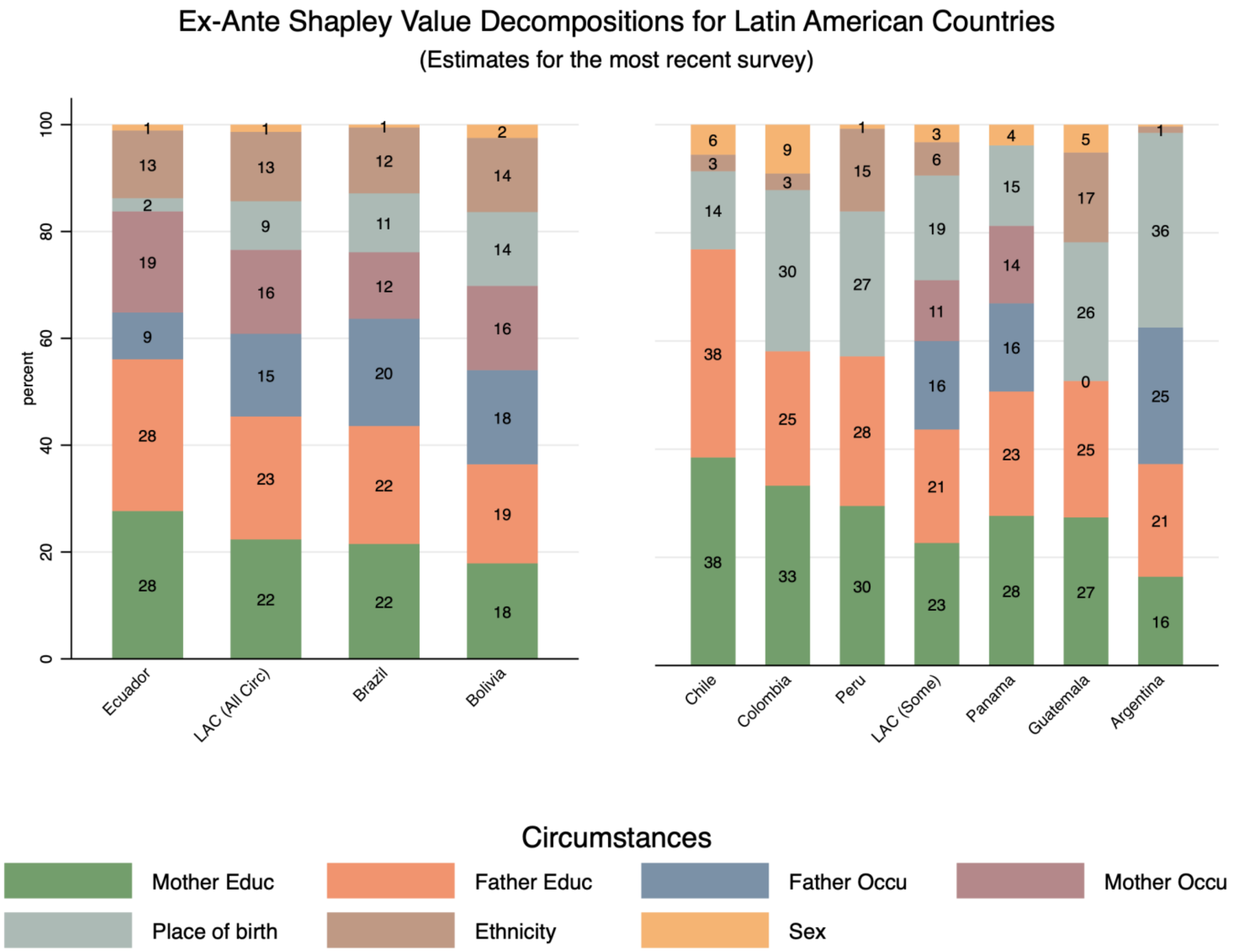
There is a level difference between MLD and Gini-based estimates, but almost perfectly correlated



Note: MLD estimates in the paper and appendix slides.

5. Ex-ante Inequality of Opportunity

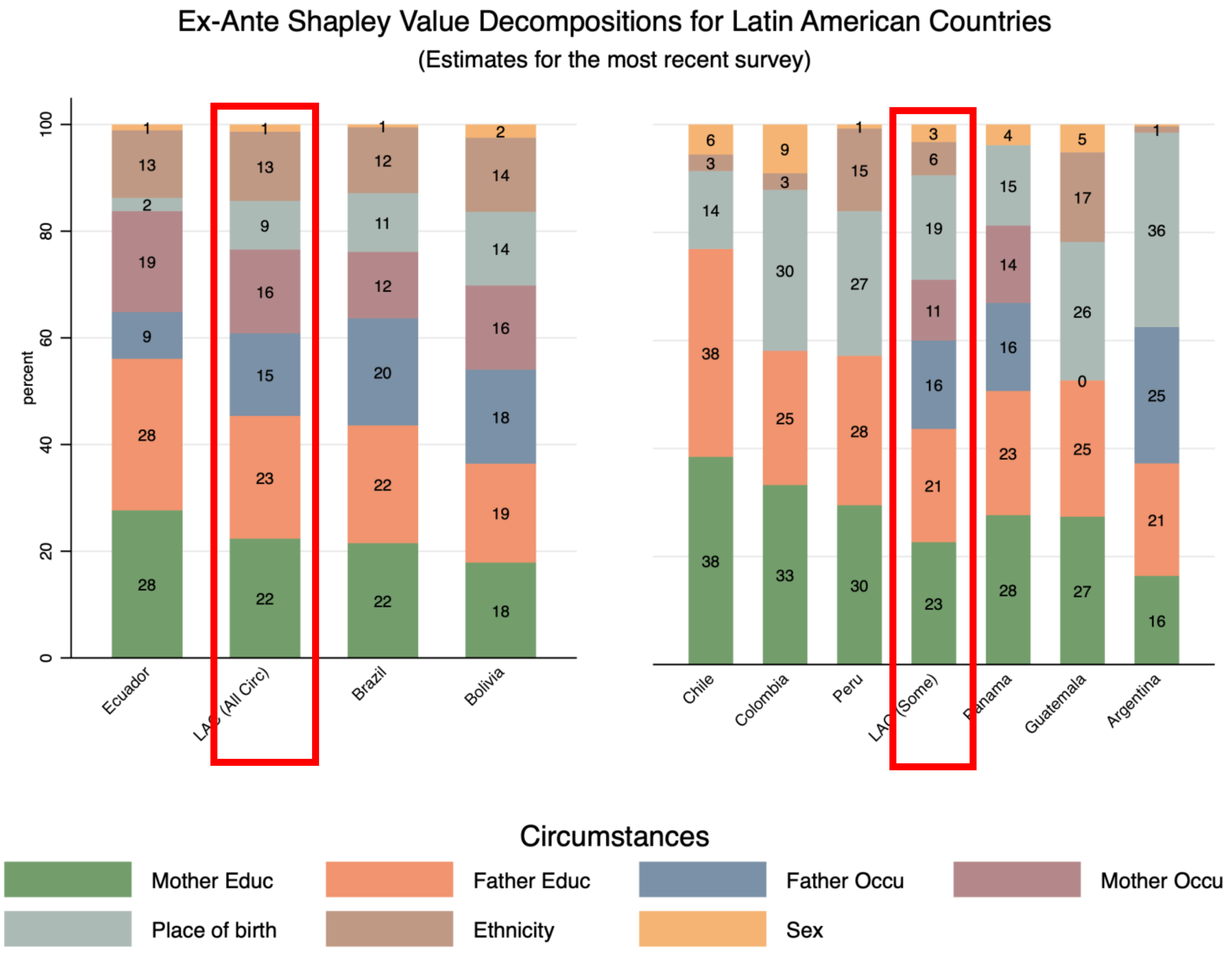
The relative importance of individual circumstances can be estimated by a Shapley decomposition



Parental background is hugely important; interesting variation in other variables

5. Ex-ante Inequality of Opportunity

The relative importance of individual circumstances can be estimated by a Shapley decomposition



Parental background is hugely important; interesting variation in other variables

6. Ex-post Inequality of Opportunity

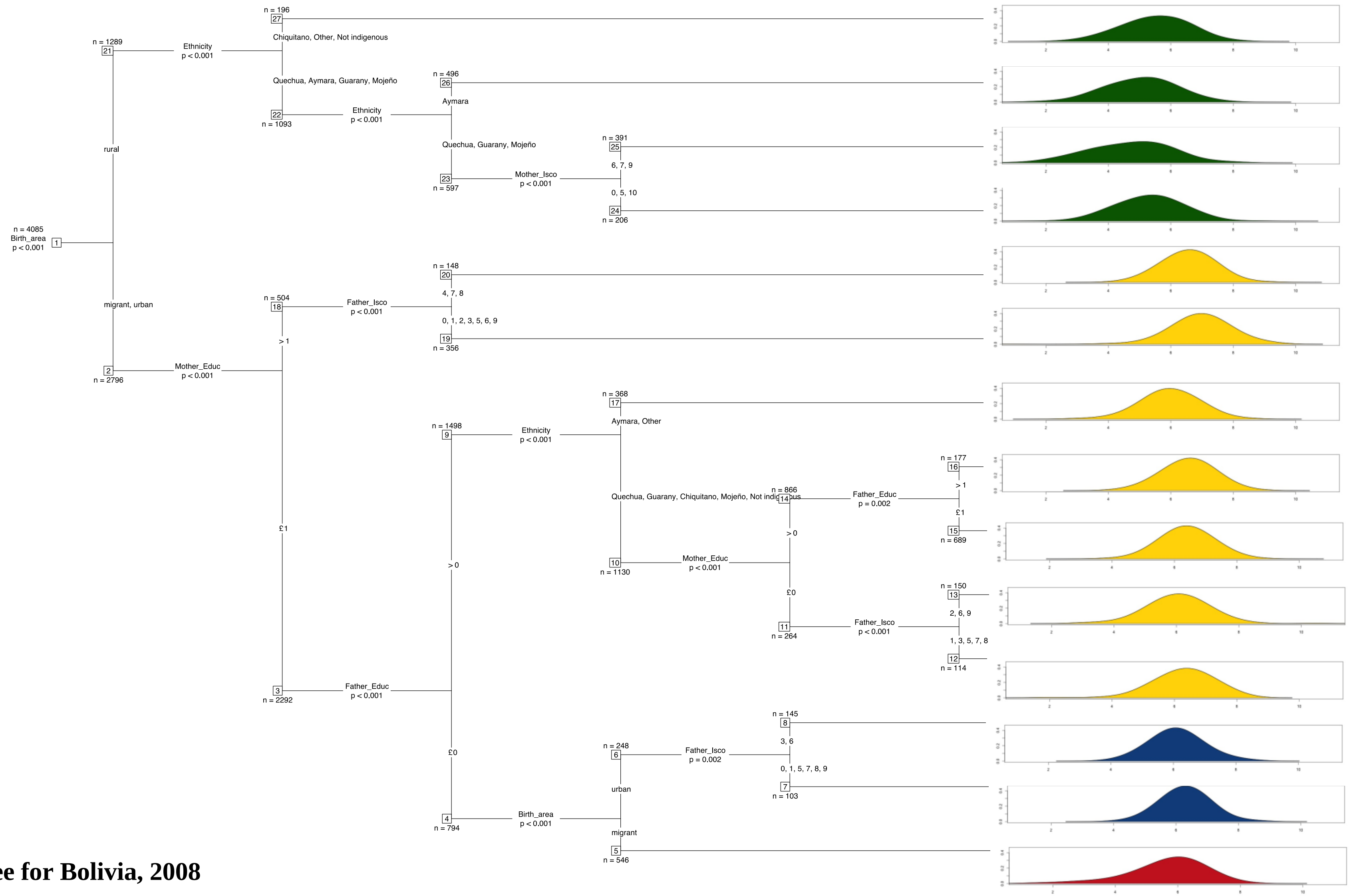


Figure 2: Transformation tree for Bolivia, 2008

6. Ex-post Inequality of Opportunity

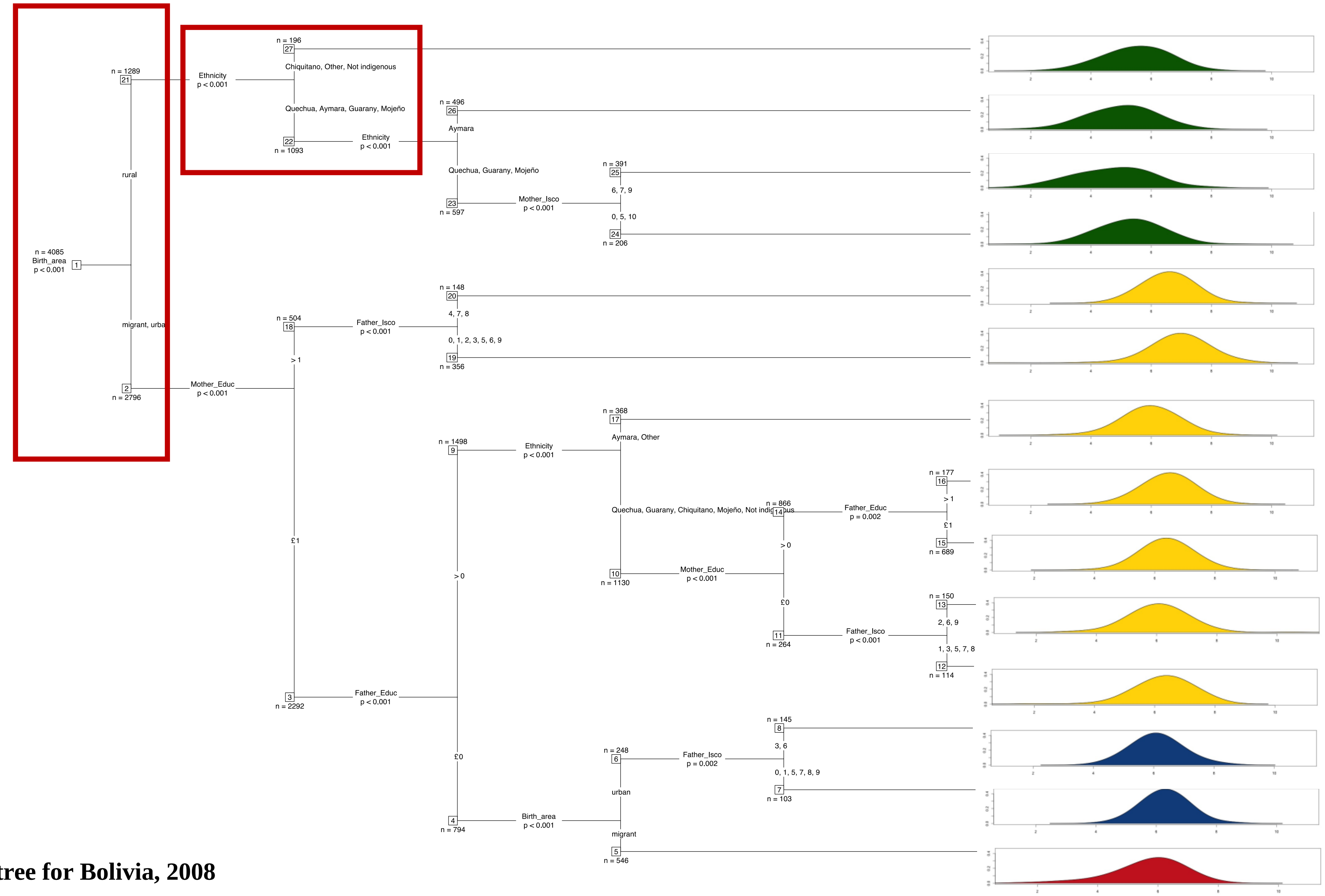


Figure 2: Transformation tree for Bolivia, 2008

6. Ex-post Inequality of Opportunity

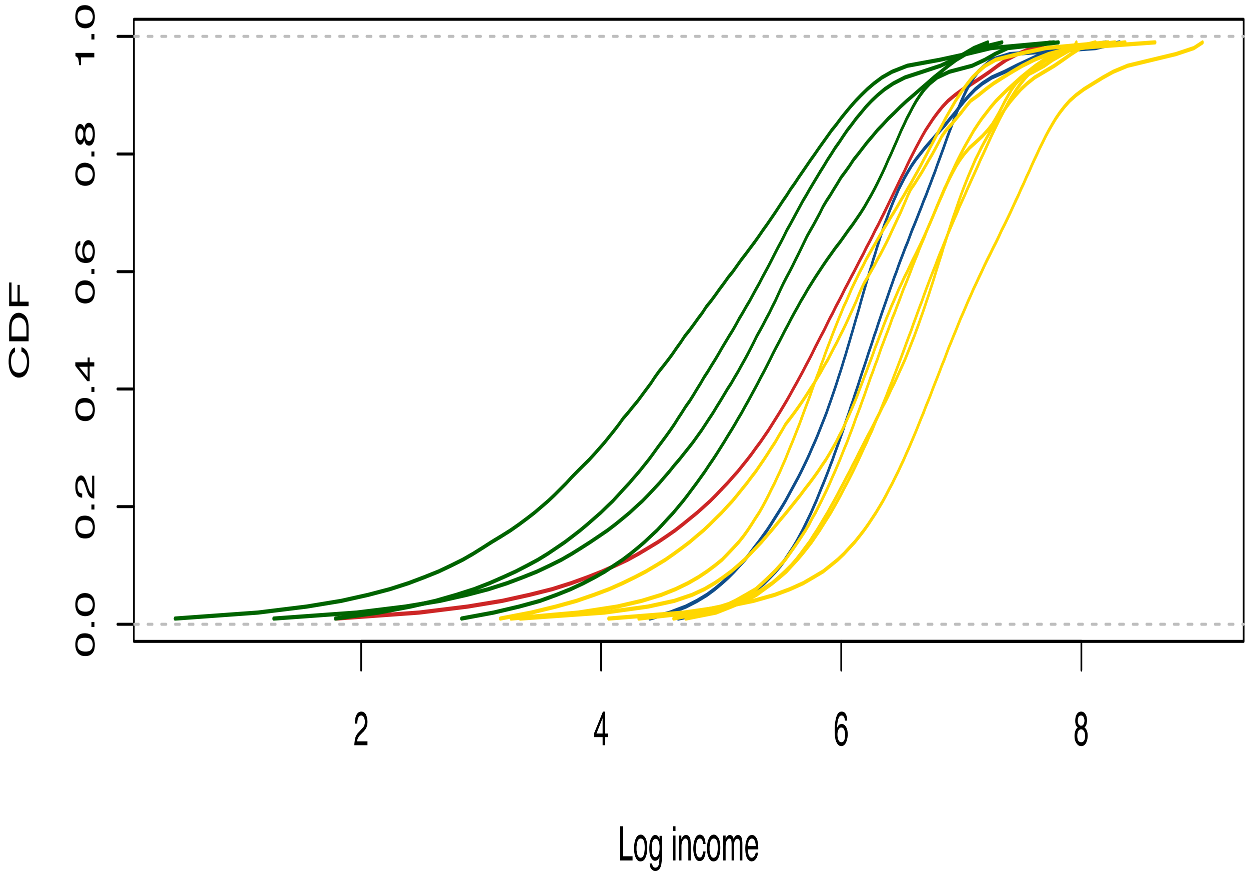
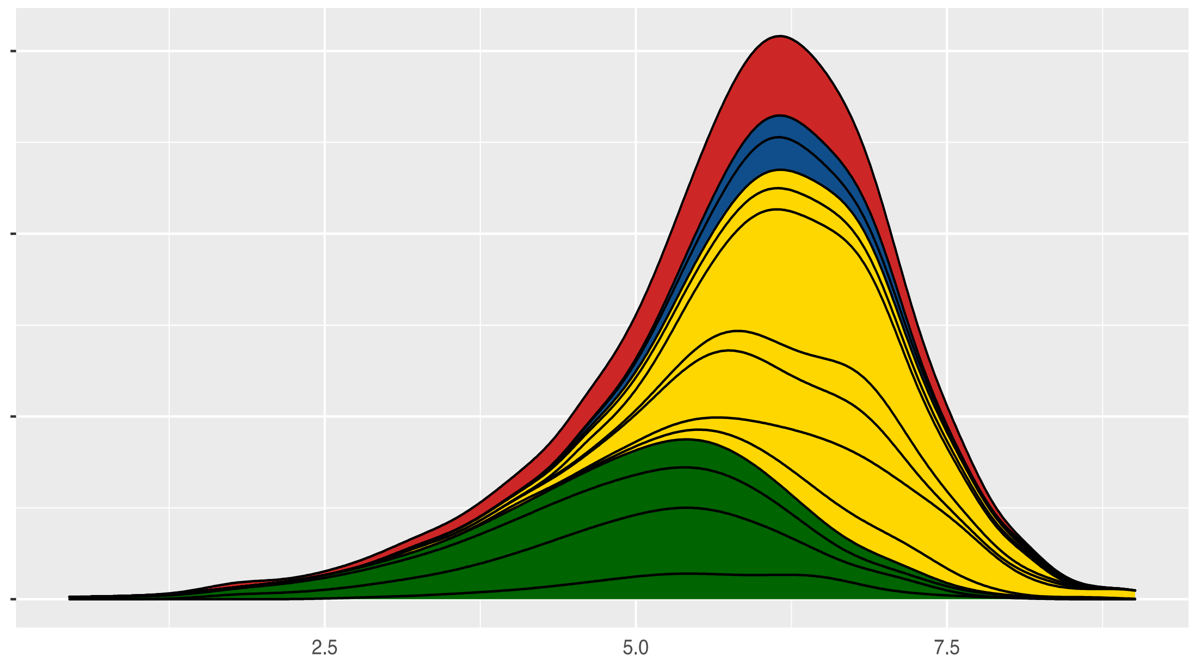


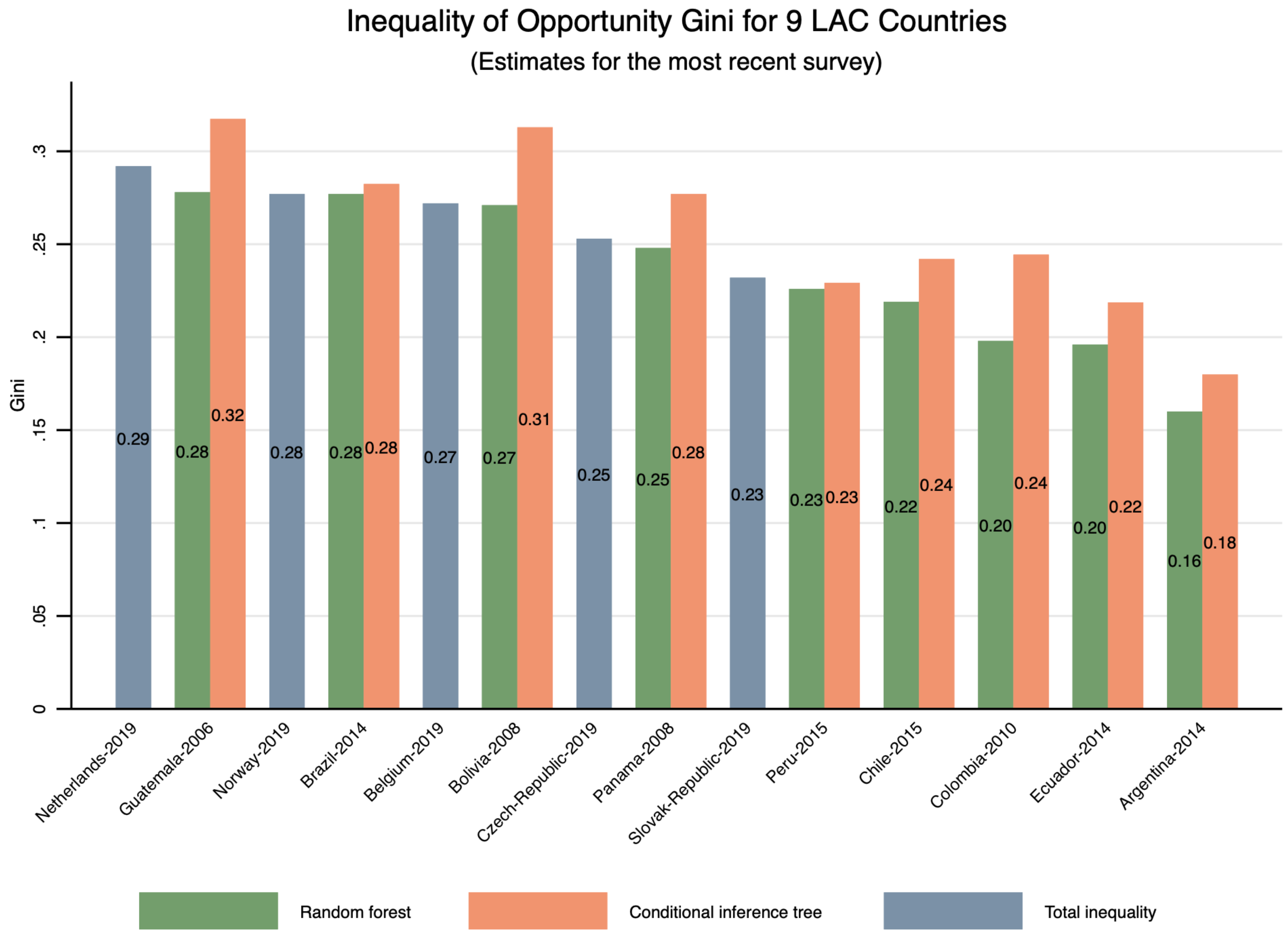
Figure 4: The income distribution in Bolivia as mixture of 14 type specific distributions



Note: Colors refers to area of birth: red=migrant, blue= urban, gold=migrant and urban, green=rural

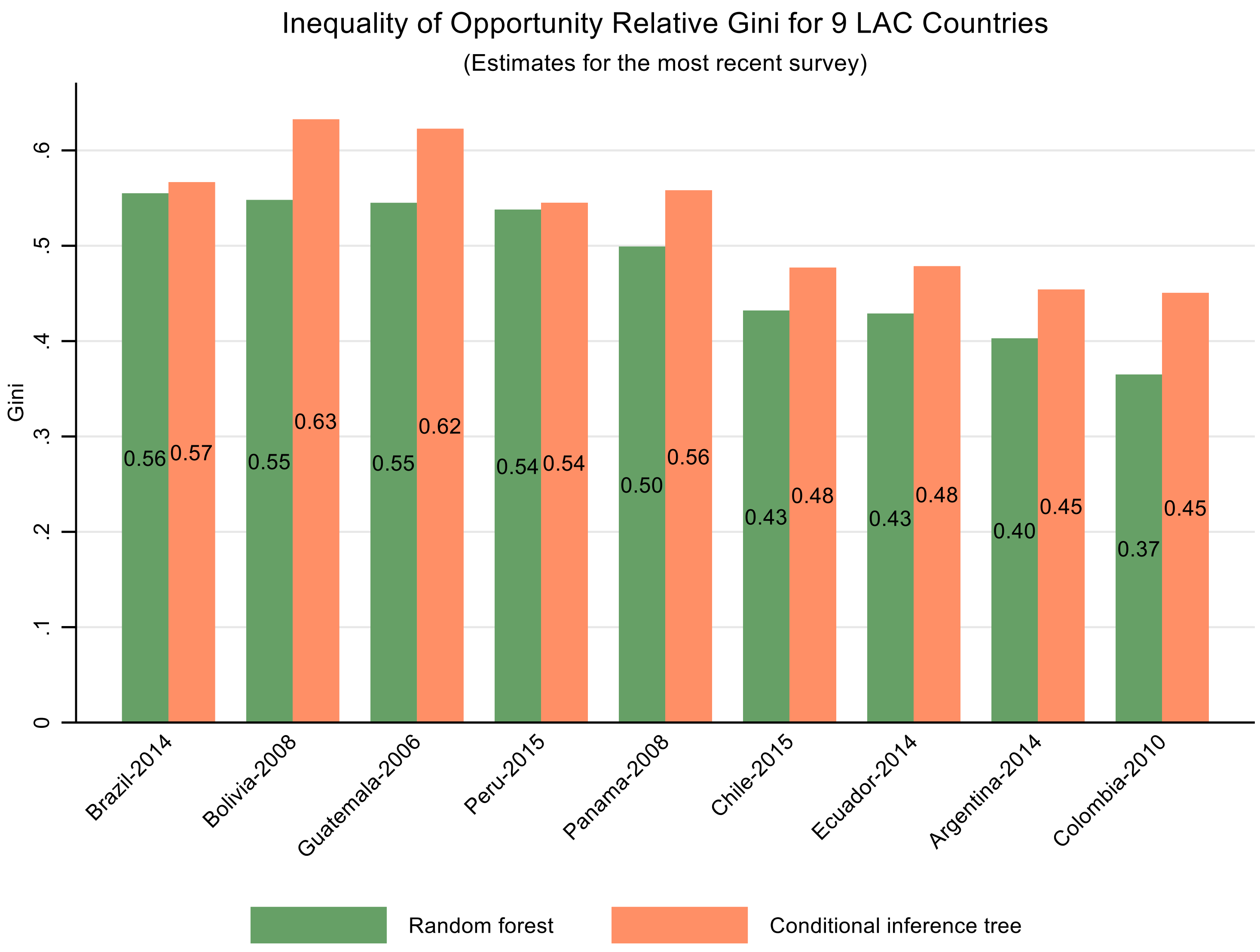
Note: Colors refers to area of birth: red=migrant, blue= urban, gold=migrant and urban, green=rural

6. Ex-post Inequality of Opportunity



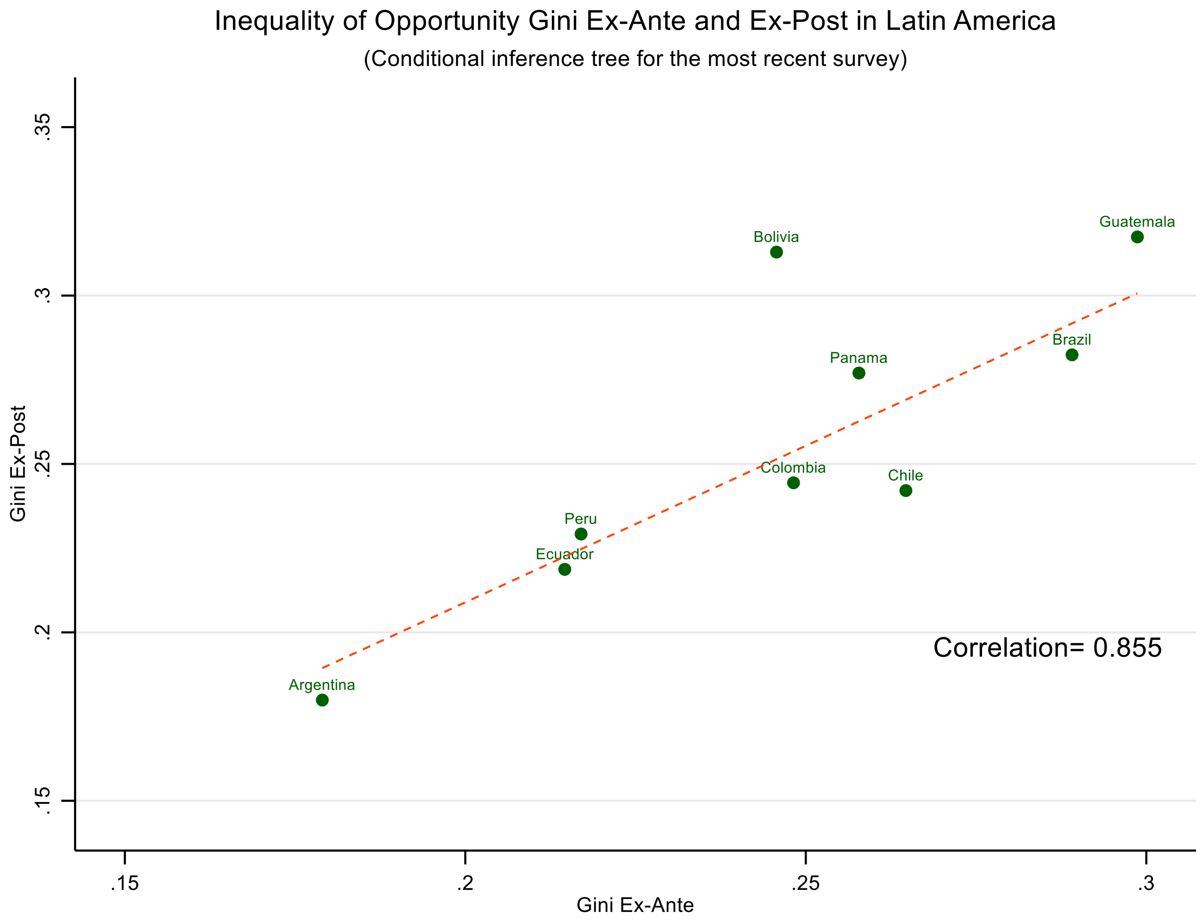
Source: World Development Indicators online (21 August 2022) for Belgium, Czech Republic, Netherlands, Norway, Slovak Republic.

6. Ex-post Inequality of Opportunity



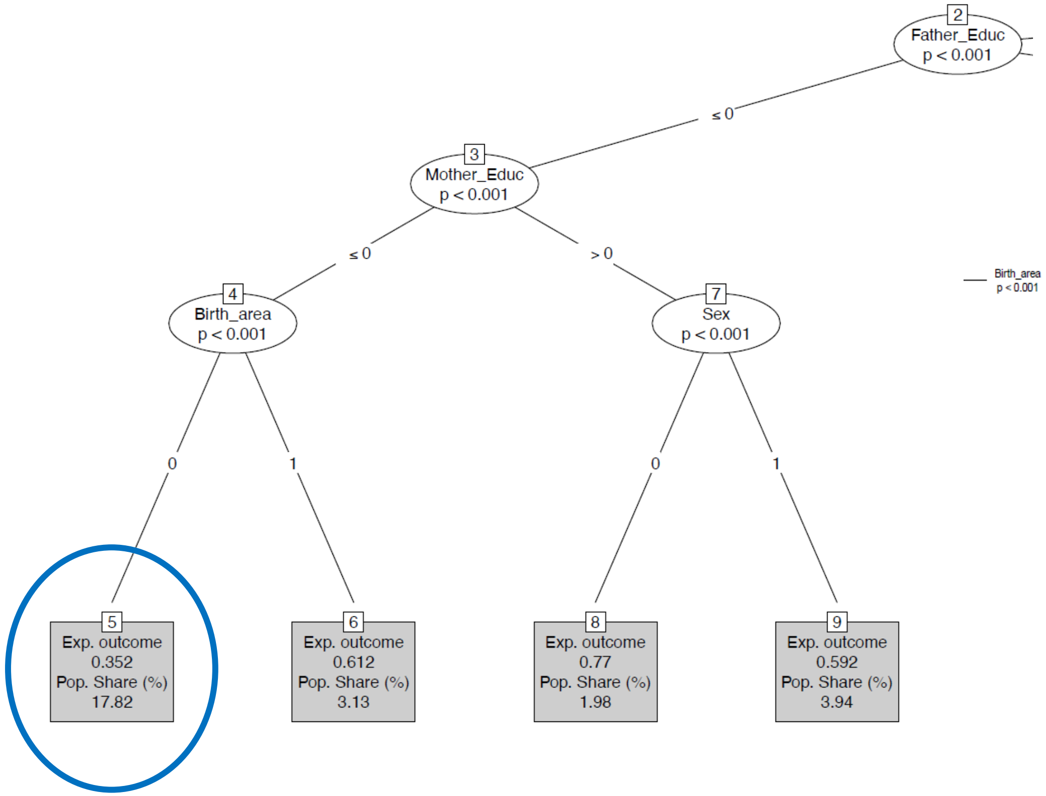
7. Comparisons and conclusions

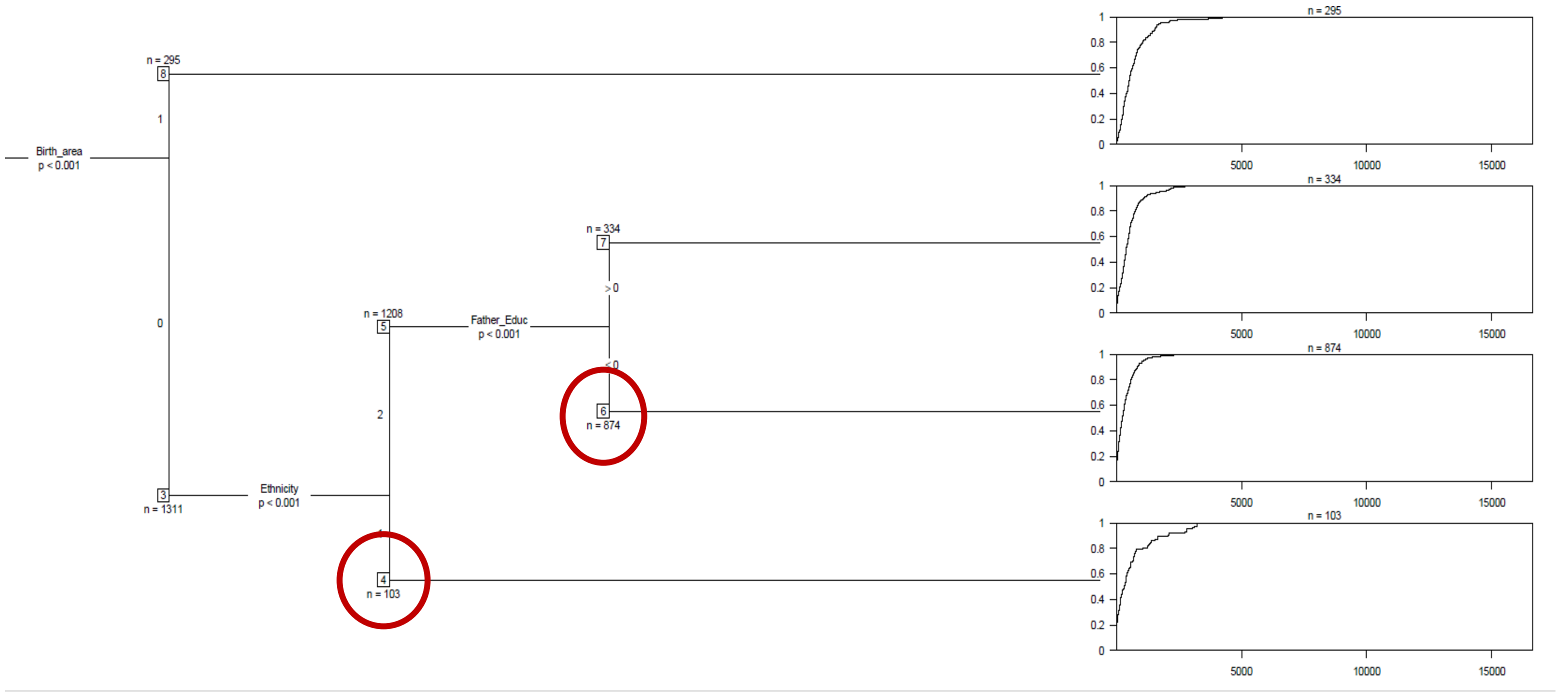
Ex-ante and ex-post measures of IOp are closely – but not perfectly – correlated in our sample.
This may reflect estimator variance to some degree, but it clearly also reflects conceptual differences.



7. Comparisons and conclusions

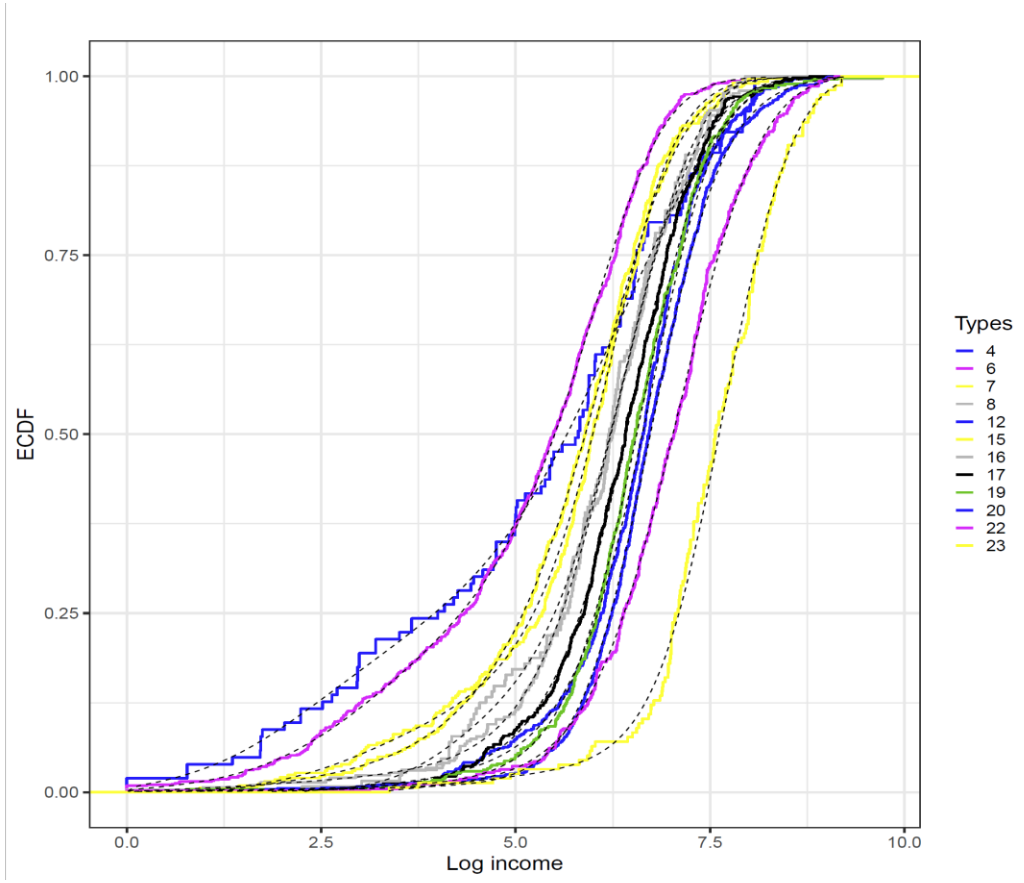
- Example of differences in the type partition between the ex-ante and ex-post approaches, given sensitivity to higher moments: Tree excerpts from Panama (2003).

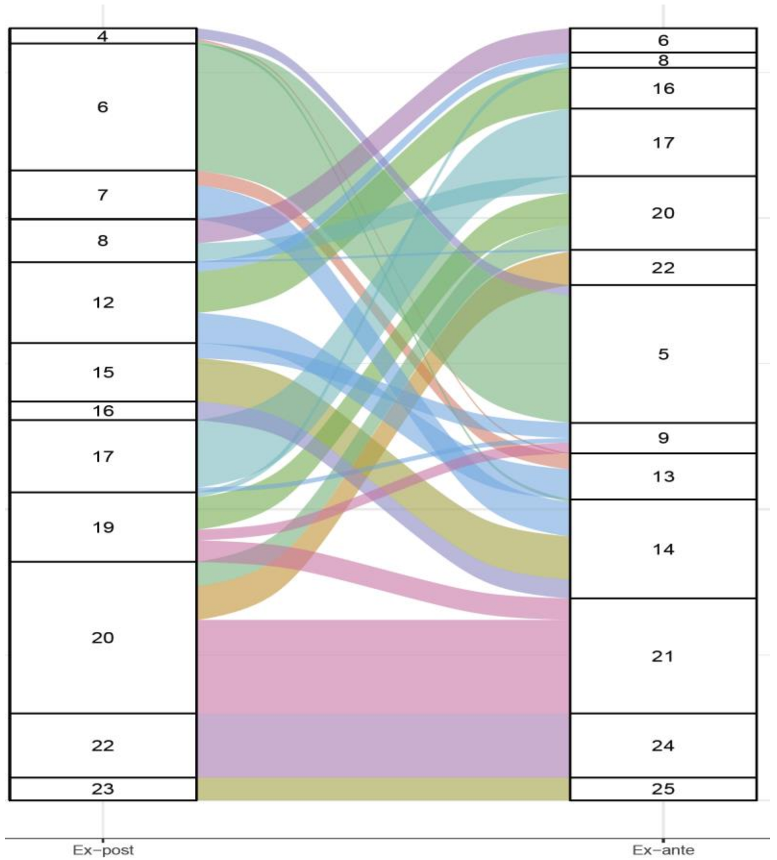




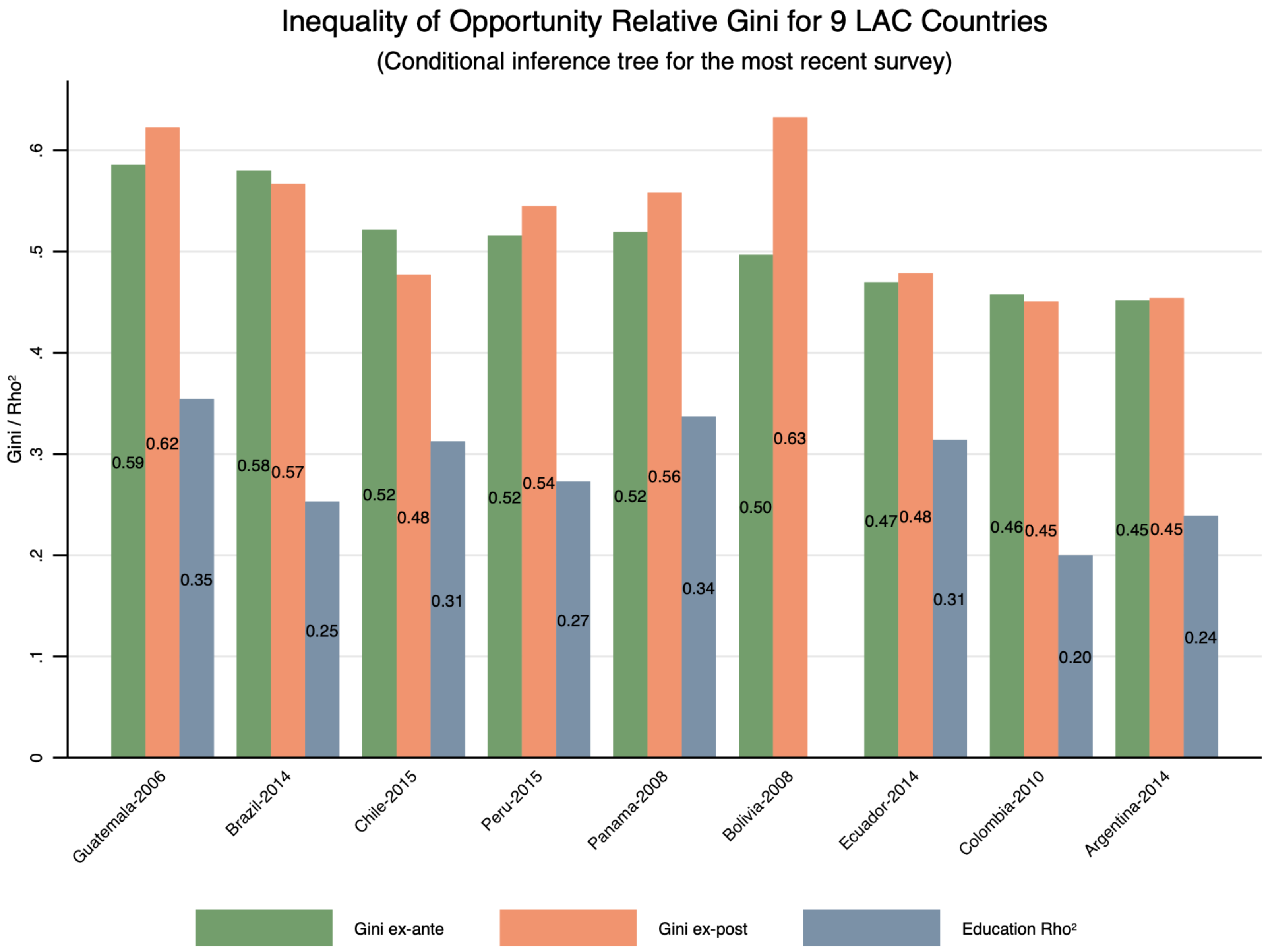
7. Comparisons and conclusions

- Little difference between the means of the two poorest types in TrT, but bigger differences in higher moments. 100% of Type 6 and 70% of type 4 are in EA Type 5.





7. Comparisons and conclusions



7. Comparisons and conclusions

1. **Socioeconomic advantage, as measured by income or education, is highly persistent in Latin America and the Caribbean**

   - In our sample, correlation coefficients for years of schooling range from 0.45 to 0.60.
   - The Opportunity Gini for income ranges between 0.18 and 0.30 (ex-ante trees) and 0.18 and 0.32 (ex-post trees) - higher than overall income inequality in some countries.
   - As a share of total income inequality, the OpGini ranges between 45% and 59% (ex-ante trees) and 45% and 63% (ex-post trees).
   - Descriptively, parental education and occupation are the most salient circumstances, at least in the ex-ante analysis.
   - CI and transformation trees are informative of the structure of inequality of opportunity in LAC countries, and reveal interesting cross-country differences in the role of, say, ethnicity and birthplace
   - Share of current variation “explained” by inherited circumstances obtained from this new approach are considerably higher than, say, from IGM in education.

Many thanks.

Muchas gracias.