CEQ INSTITUTE
COMMITMENT TO EQUITY
Tulane University

Including the Rich in Income Inequality Measures:

An Assessment of Correction Approaches

Nora Lustig
Samuel Z. Stone Professor
Founding Director of CEQ Institute
Tulane University

III-LIS Comparative Economic Inequality Conference
February 27, 2025
LIS, Luxembourg

Based on (cite as):

Nora Lustig and Andrea Vigorito. (2025). Including the Rich in Income Inequality Measures: An Assessment of Correction Approaches. Forthcoming.

Outline

- Introduction
- Correction approaches
- Sensitivity of Results
- Choosing the Correction Approach
- Conclusions

Introduction

The Issue with Household Surveys

- Personal income inequality is traditionally measured using household surveys

- **Limitation:** HH surveys often fail to accurately capture income in the upper tail of the distribution, especially income derived from capital **=> The "missing rich" problem**

- **Causes:** sampling errors, coverage errors, unit and item nonresponse, underreporting and preprocessing practices (e.g., top coding)

- **Consequences:** biased survey-based income distribution and inequality indicators; levels and trends can be affected

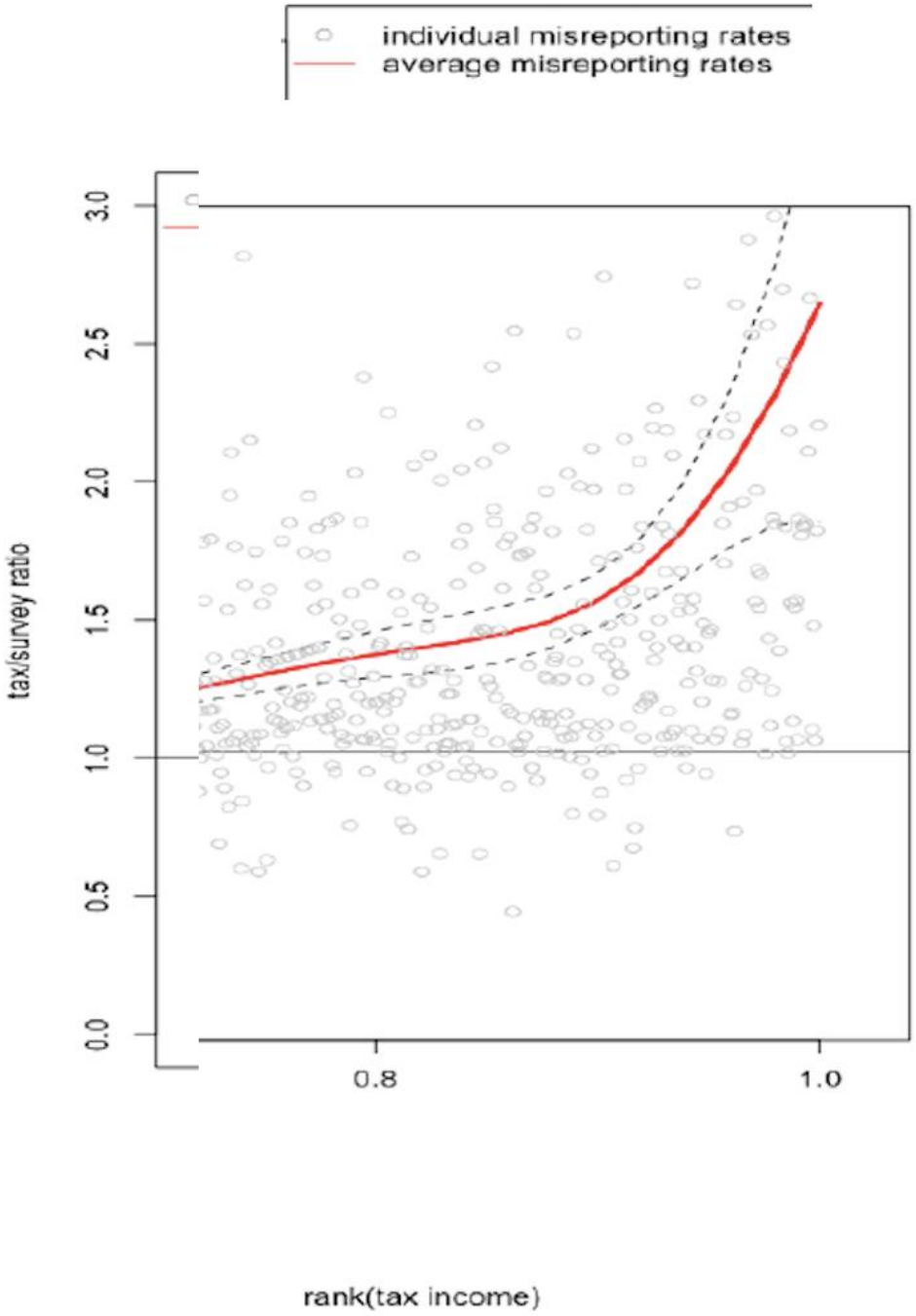
  - Beyond measuring inequality, inaccurate inference of determinants of inequality and the relationship between inequality and, for instance, growth

Evidence the Rich are Missing in HH Surveys

- By inspection
- Comparison with external data (e.g., tax records)
- Nonparticipation Rates in Surveys
- Evidence from Linked Data

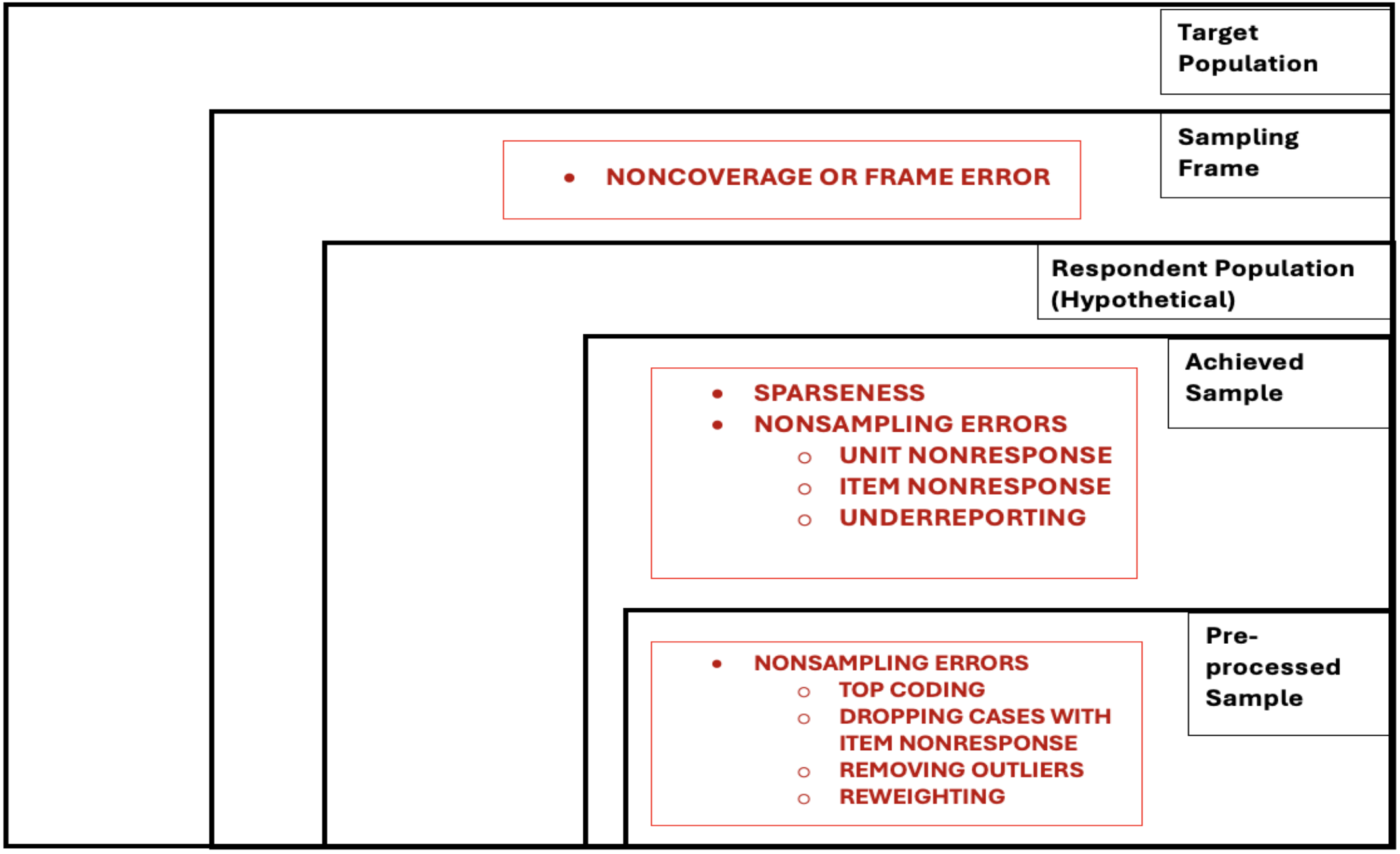
Top incomes in surveys: high-paying managers, lawyers, doctors, engineers

Not in hh surveys....



- Underreporting rises with income
- Evidence from linked data for Uruguay

Flachaire, Lustig and Vigorito (2023)



Source: adapted from Figure 1, Lustig and Vigorito (2025).

Searching for Solutions

- Renewed interest in economic inequality led to a corresponding interest in addressing the 'missing rich' problem (Atkinson and Piketty, 2007; Milanovic, 2023)

➢ The goal of the approaches is to generate inequality measures that are a more accurate estimate of the actual (unobserved true) distribution of income

- Approaches proposed in the literature fall into three main strands:

1. Correcting household surveys (references shown later)
2. Relying on external data such as tax records (Atkinson and Piketty, 2010, 2011)
3. Distributional National Accounts (Zwijnenburg, 2019; Blanchet et al., 2024; WID.World/PSE; DNA/OECD et al.)

Searching for Solutions

- Focus here is on approaches to correct HH surveys
- The goal of these approaches:

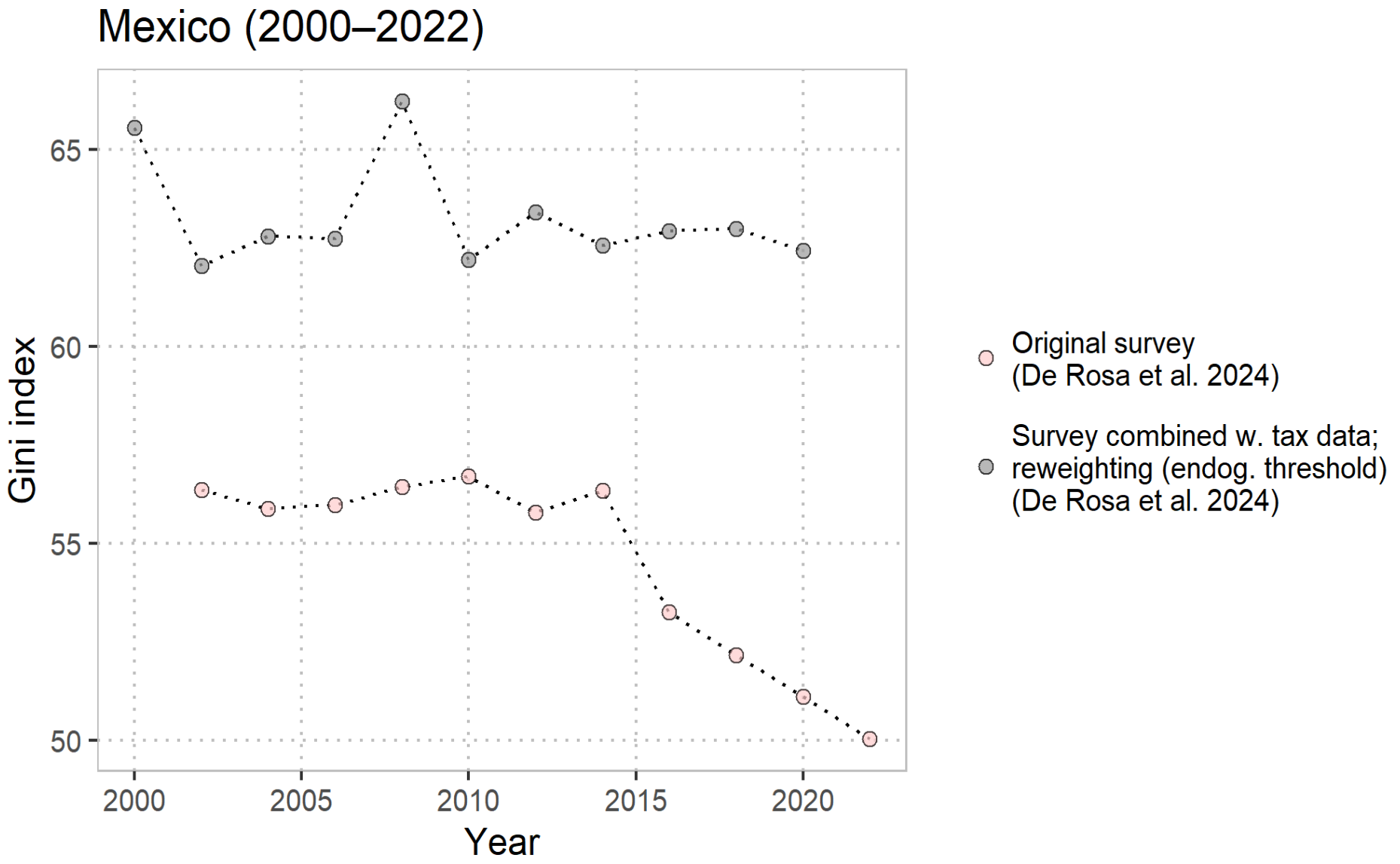
➢ Transform household survey so that corrected distribution is a more accurate representation of actual (unobserved true) distribution of income

➢ Identified **TWENTY-TWO** distinct approaches that have been applied in practice

The Issue With Correction Approaches

-Unsurprisingly, inequality levels, and sometimes trends, change after correction

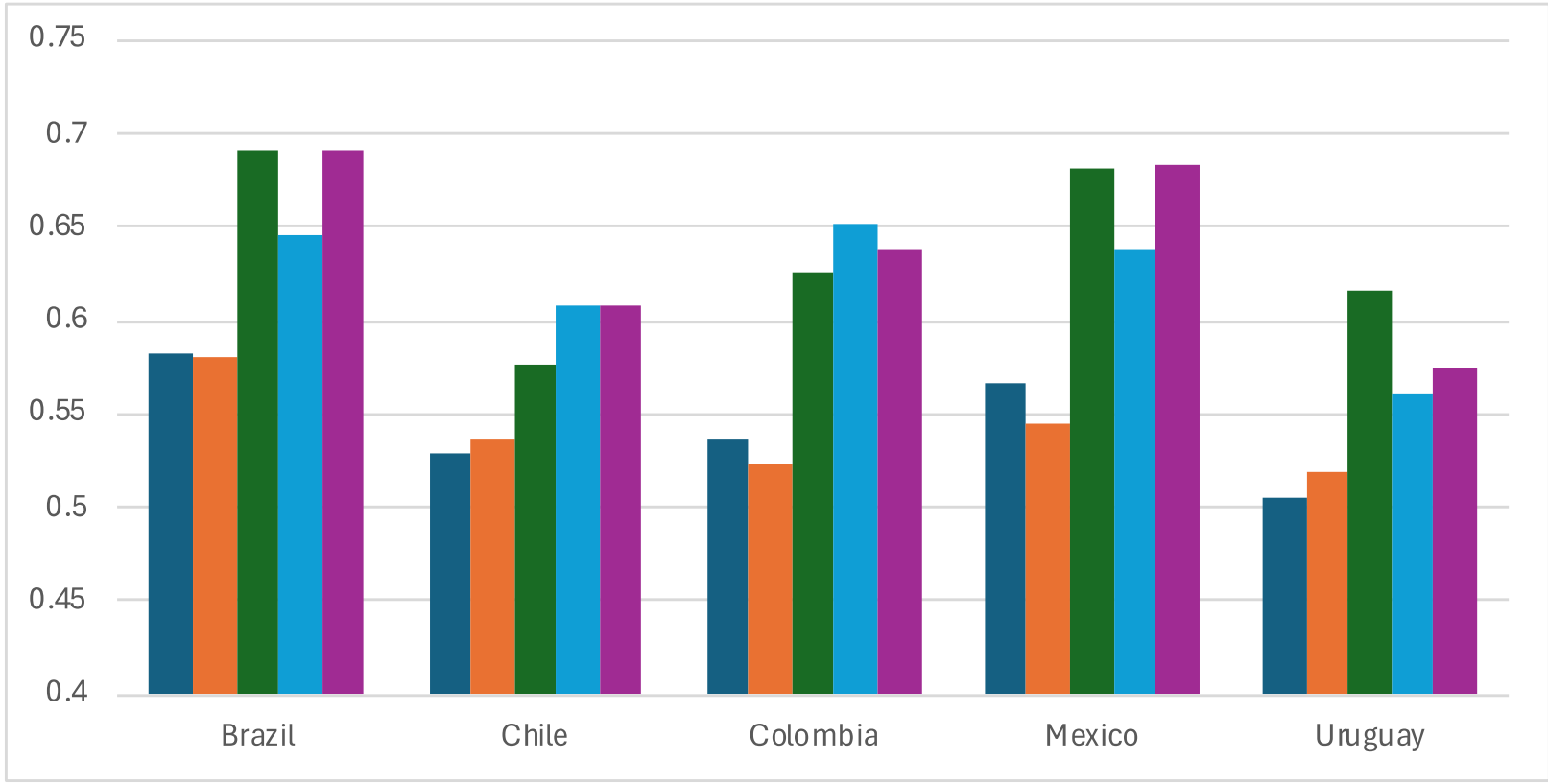
-In general, higher; but remember that corrected inequality can be lower (Deaton, 2005)



The Issue With Correction Approaches

- In addition, evidence suggests that inequality indicators are highly sensitive to the *specific* correction method
  - Applying different correction approaches to the *same* data can yield significantly different results for the *same* inequality indicator
  - Moreover, there is no consistent pattern in how different methods influence inequality estimates
  - The impact of correction methods on inequality indicators varies depending on how each method transforms the data

Gini Coefficient: Same Data, Different Correction Approaches



De Rosa, Lustig and Martinez Pabon (forthcoming)

What's To Be Done

- What criteria should be used to determine which methodological approach brings us closer to the true inequality
  - Unfortunately, there are no statistical tests or calibration mechanisms available to make this determination
  - In practice, researchers rely on assumptions, often made implicitly, and ad hoc considerations to decide which approach to follow
  - Broad guidelines to make decisions on the method and various aspects within each method
  - Nevertheless, it may not be possible to generate a single, definitive inequality indicator

Correction Approaches

Towards a Taxonomy of Correction Approaches

- Goal: transform household survey so that the corrected distribution is a more accurate estimate of the actual (unobserved true) distribution of income

- Classify correction approaches: according to source of data and method (Hlasny and Verme, 2018a; Lustig, 2019)

- Source of the data: within-survey or survey with external data (tax records, social security registries, or National Accounts)

- Correction method: replacing, reweighting, and reweighting and replacing

- Within each one, there are a number of submethods

- Each approach/method/submethod has key underlying assumptions

Correcting Household Surveys: Approaches

- **Replacing**
  - Original income observations at the top in survey are replaced with a (presumably) more accurate estimate of the upper tail
  - Original weights assigned to the top as a whole and to the rest of the distribution remain intact
- **Reweighting**
  - Adjusts the weights assigned to different income groups in the survey to (in principle) better reflect the true representation of high-income individuals
  - Original income observations remain intact
- **Reweighting and replacing**: modifies both the income observations and the weights

• A variety of submethods: parametric and nonparametric replacing, model-based reweighting, poststratification

Replacing vs. Reweighting: the Key Distinction

The main distinction between replacing and reweighting :

- Whether the population shares that define the top and the rest, the original weights and the original incomes remain the same or change in the corrected distribution

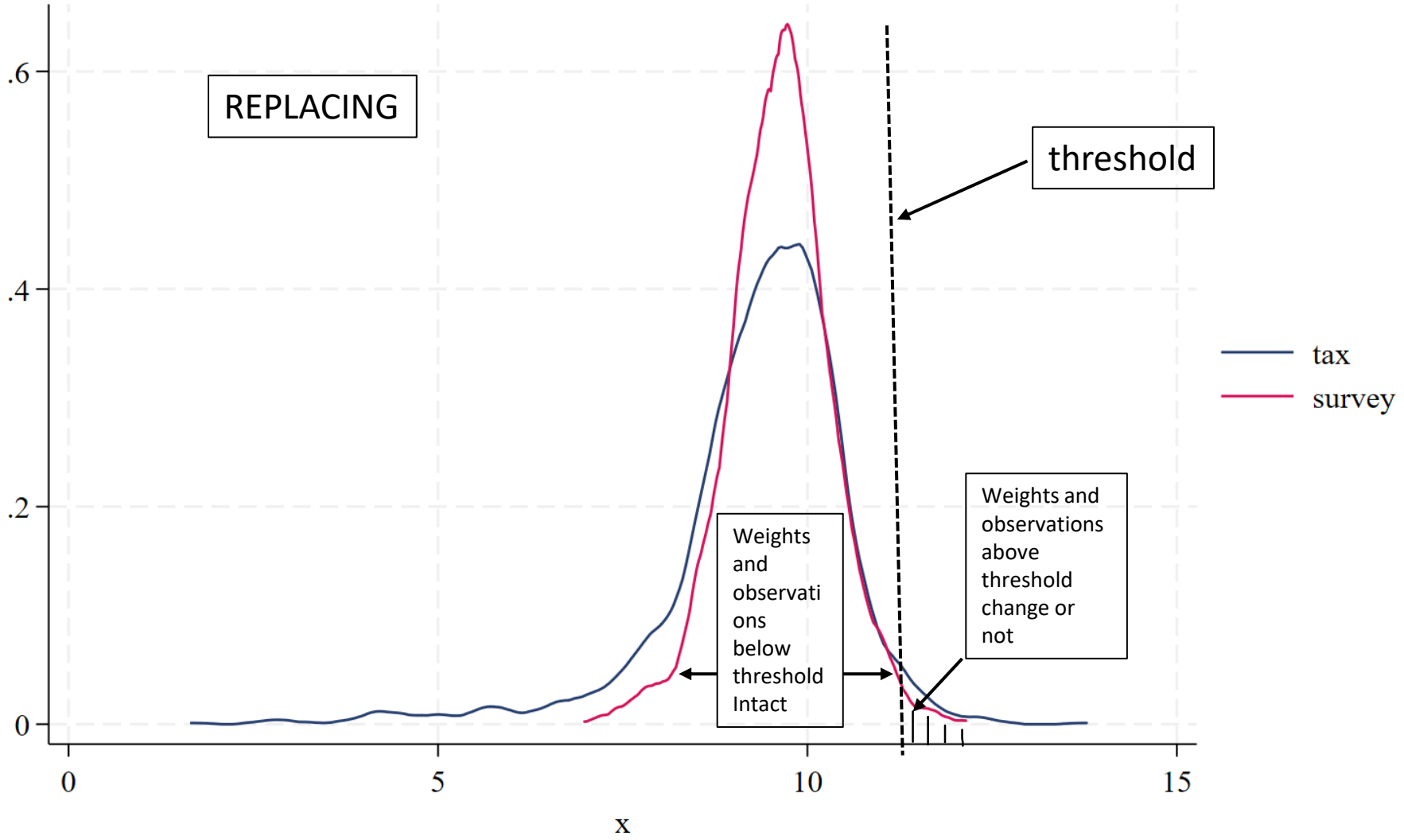
- In replacing,
  - The population share above the threshold that defines the upper tail that needs correction remains the same
  - The original weights remain intact except *within* the top in some methods
  - The incomes above threshold change while remain intact at and below the threshold; except in reweighting within the top
- In reweighting, the weights change, including the share of the upper tail, and incomes remain intact
- When reweighting and replacing are combined, the original incomes and the weights change

Method	x = incomes	w = weights	Beta = Share of population in upper tail
Replacing	change above the threshold but remain intact below	remain intact except in some methods weights change *within the top*	remains the same after correction
Reweighting	remain intact	change throughout	changes after the correction
Replacing&Reweighting	change above the threshold but remain intact below	change throughout	changes after the correction

Within-survey

or

Combining survey and external data

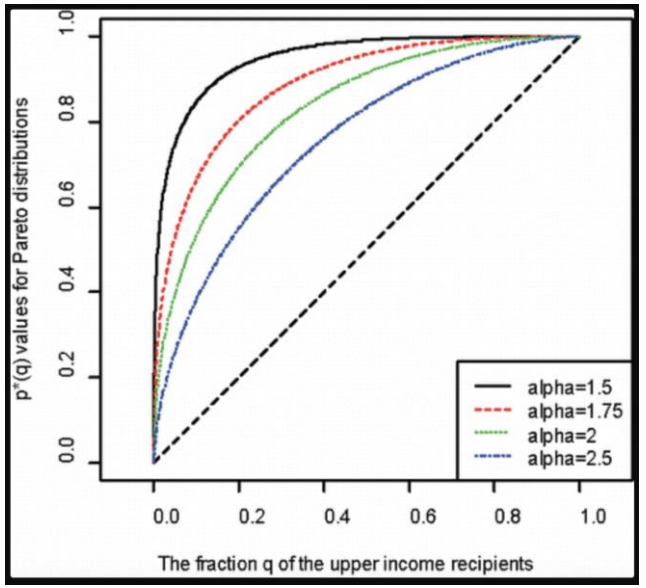

REPLACING
threshold
tax
survey
Weights and observations below threshold Intact
Weights and observations above threshold change or not
.6
.4
.2
0
0
5
10
15
x

Replacing

- Select threshold
- Select data
  - Survey
  - Survey and tax records
  - Survey and National Accts.
- Select method
  - Semiparametric
    - Select parametric function
    - Select estimation model
  - Nonparametric
    - Within-survey imputation
    - Survey and external data
      - Rescaling
      - Statistical matching
      - Reweighting within the top

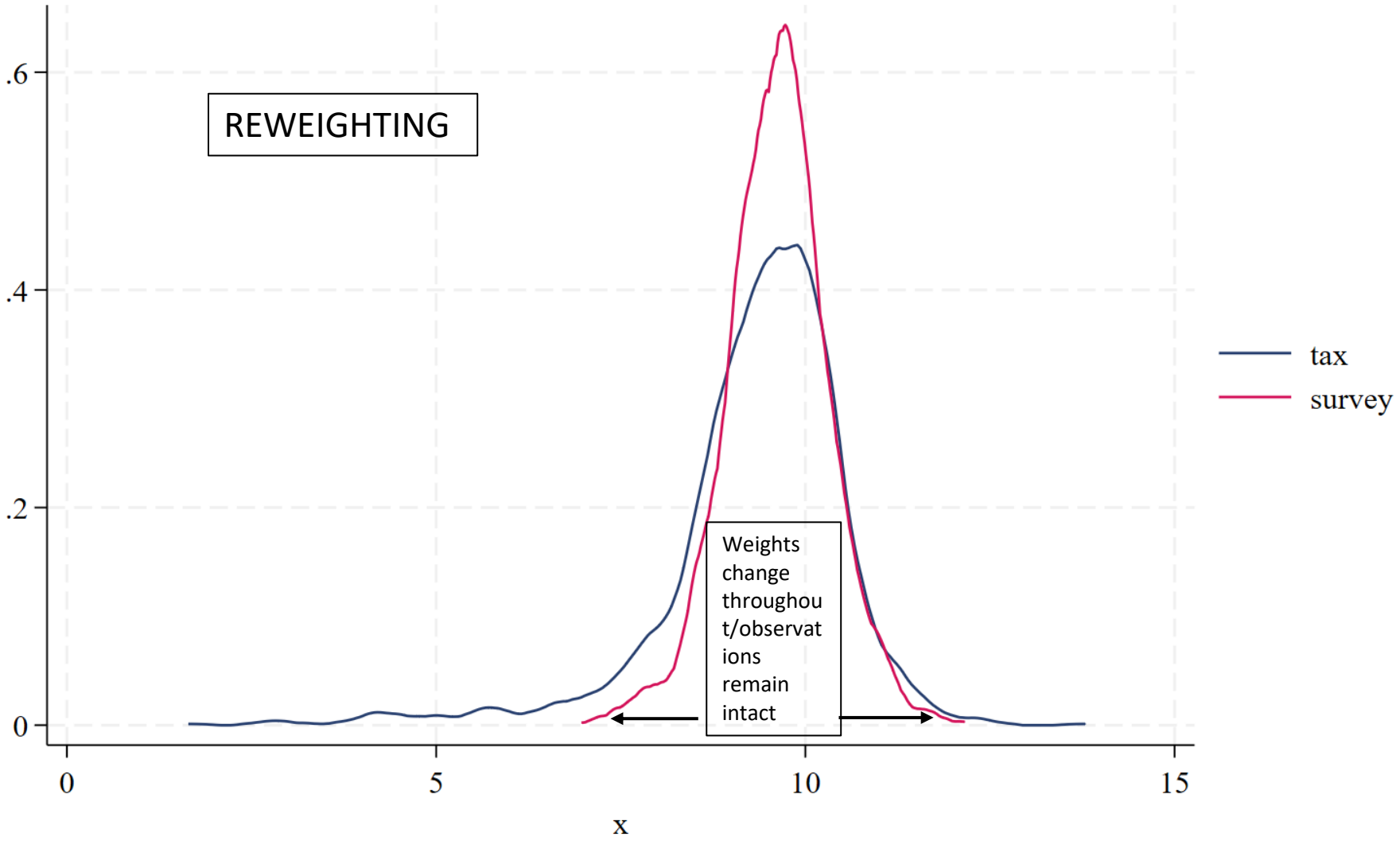
APPROACH	REPLACING				
	Within-survey		Survey and External Data		
METHOD	Semiparametric	Nonparametric Imputation	Semiparametric	Nonparametric	
			External Income Microdata; Regression-based Income Prediction; Income Totals (Nat Accts)	Rescaling; Statistical Matching; Replacing w/External Data	Reweighting Top with External Income Microdata
Assumes Common Support	No	Yes	No		Yes
Weight of the Upper Tail and the Rest Intact	Yes				
Weights within Rest of Distribution Intact	Yes				
Observations (incomes) within Upper Tail Intact	No				Yes
Absolute Poverty Indicators Intact	Yes				

REPLACING BY A PARAMETRIC DISTRIBUTION



TOP PERCENT IN
SURVEY ARE TOSSED OUT

Method: Replaces the top percent (1, 5, 10 percent for ex) of the distribution by parametric distribution (e.g., Pareto) estimated with observations in survey or external data (tax records)


REWEIGHTING
.6
.4
.2
0
0
5
10
15
x
Weights change throughout/observations remain intact
tax
survey

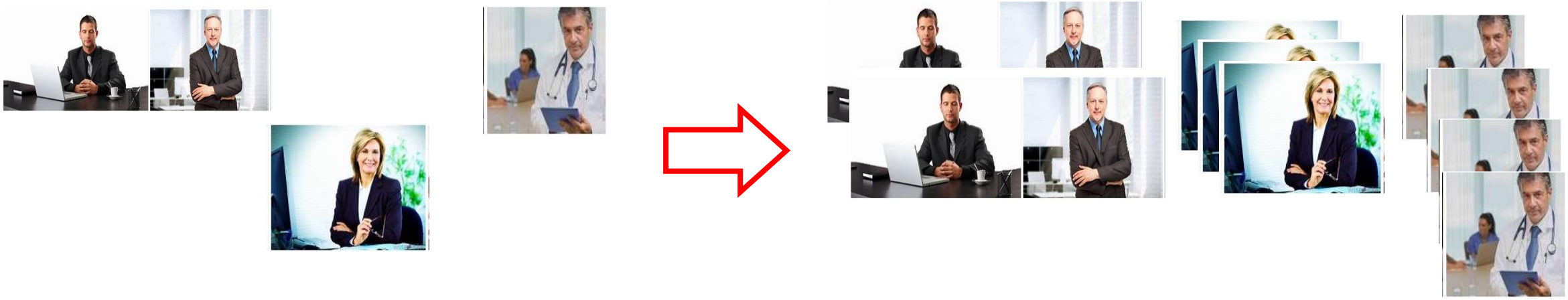
Reweighting

- Threshold
  - No preselection
  - Preselected
- Select data
  - Survey
  - Survey and tax records
  - Survey and National Accts.
- Select method
  - Within-survey
    - Model Weight Adjustment
  - Survey & external data
    - Poststratification
    - Increase weight at top; downweigh uniformly the rest

APPROACH	REWEIGHTING			
	Within-survey		Survey and External Data	
METHOD	Weighting Class Adjustment	Model Weight Adjustment	Poststratification	Reweighting Top and Uniform Downweighting of Rest
Assumes Common Support	Yes			
Weight of the Upper Tail and the Rest Intact				No
Weights within Rest of Distribution Intact				No
Observations (incomes) within Upper Tail Intact	Yes			
Absolute Poverty Indicators Intact				No

REWEIGHTING

New weights based on, for ex, nonresponse adjustment factor or poststratification weights (Little & Rubin, 2014; Biemer & Christ, 2008)



Weights of these observations were reduced

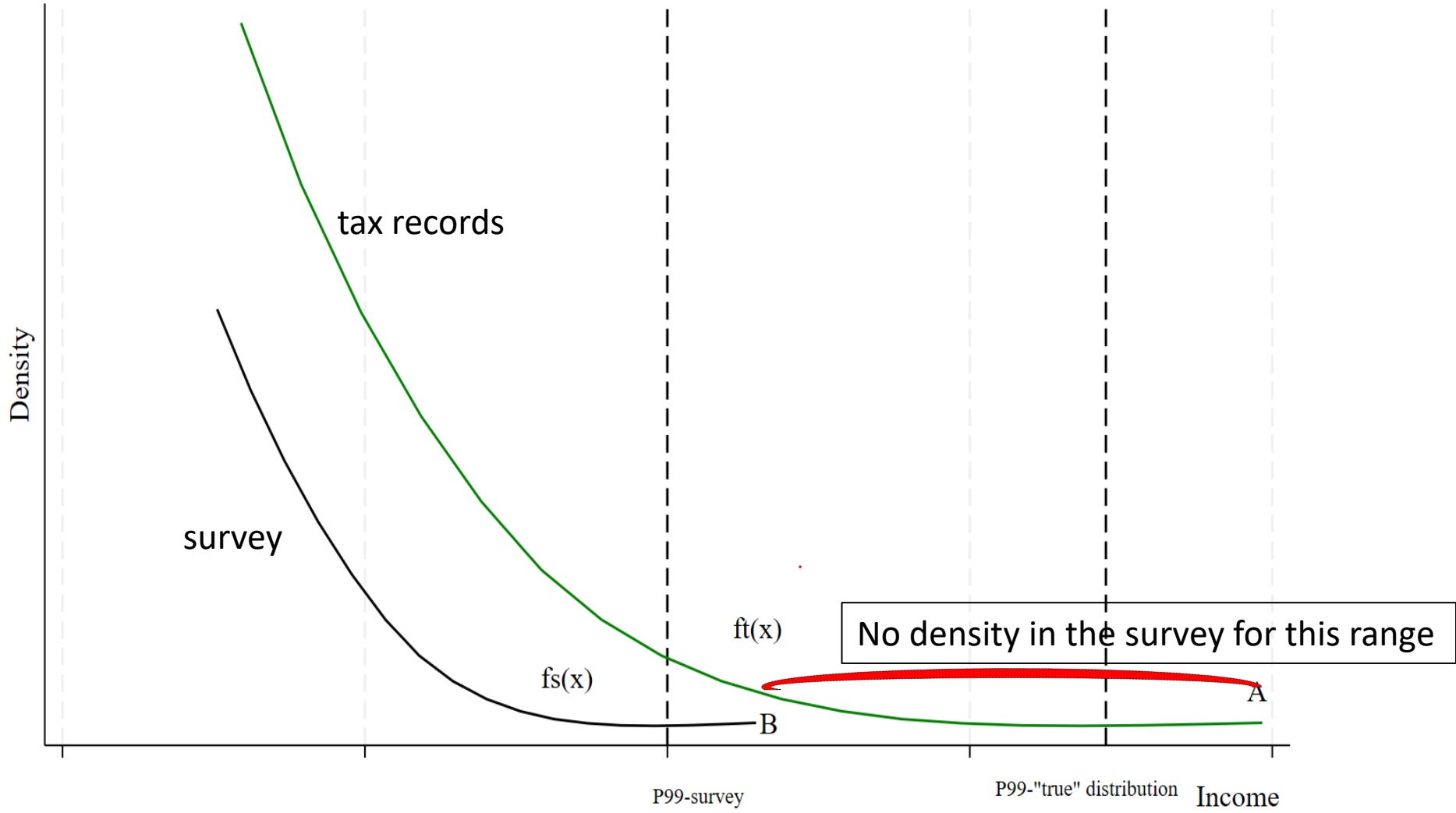
OBSERVATIONS IN SURVEY ARE REWEIGHTED TO TAKE INTO ACCOUNT NONCOVERAGE, UNIT AND ITEM NONRESPONSE; NEW WEIGHTS WITHIN-SURVEY OR POSTSTRATIFICATION METHODS

- Replacing as a correction method can work if the presumption is that the weights—except perhaps within the top-- in the original uncorrected survey adequately represent the target population
- In other words, replacing assumes that undercoverage or unit or item nonresponse or underreporting are *within* the upper tail
- If, however, there is evidence that, for example, nonparticipation in survey rises with income, then replacing will not fix this => reweighting

- Reweighting allows one to retain both the statistical integrity of the survey design (with implications for statistical inference) and for micro-data files in distributional analysis (Ravallion, 2022)
- With replacing, this is not possible in general except for some imputation approaches
- However, in reweighting an important assumption is that the specific sample drawn includes at least some rich respondents; that is, there is common support between survey and target population
- Since it affects the weights and not the incomes, reweighting as a correction method can work properly under the assumption that there is common support between the sample and the target population

Absence of Common Support

- In the absence of common support, correction must use replacing method and possibly combine survey with external data such as tax records or National Accounts



Source: Figure 2, Lustig and Vigorito (2025).

Twenty-two Approaches in Practice

- **Replacing**
  - Semiparametric: within-survey and survey and external data (Cowell and Flachaire, 2007; Hlasny and Verme, 2021; Burkhauser et al., 2010; Alvaredo, 2011; Jenkins, 2017; Bourguignon, 2018; Hlasny, 2021)
  - Nonparametric
    - Within-survey Imputation (Hirsch and Schumacher, 2004; Bollinger and Hirsch, 2006)
    - Survey and External Data
      - Rescaling (Altimir, 1987; Bourguignon, 2018; Piketty, Yang and Zucman, 2019)
      - Statistical Matching (Bach, Corneo, and Steiner, 2009)
      - Replacing with Linked External Microdata (Flachaire et al., 2018; Bollinger et al. 2019)
      - Reweighting within the Top (Medeiros et al., 2018)

Twenty-two Approaches in Practice

- **Reweighting**
  - Within-survey
    - Weighting Class Adjustment (Harris, 1977; Atkinson and Micklewright, 1983)
    - Model Weight Adjustment (Mistiaen and Ravallion, 2003); Korinek et al., 2006; Hlasny and Verme, 2018a; 2018b; 2021)
  - Survey and External Data
    - Poststratification (Atkinson and Micklewright, 1983; Campos-Vazquez and Lustig, 2019)
    - Reweighting with Exogenous Threshold (Bourguignon, 2018; Flachaire et al., 2023)
    - Reweighting with Endogenous Threshold (Blanchet et al., 2022)

Twenty-two Approaches in Practice

- **Reweighting and Replacing**
  - Within-survey
    - Model Weight Adjustment Reweighting and Semiparametric Replacing (Hlasny and Verme, 2018a; 2021)
  - Survey and External Data
    - Reweighting with Exogenous Threshold and Semiparametric Replacing (Anand and Segal, 2015; Bourguignon, 2018)
    - Reweighting with Endogenous Threshold and Rescaling (Blanchet et al., 2022)
    - Nonparametric Replacing and Reweighting with Exogenous Threshold (Bourguignon, 2018; Burkhauser et al., 2018)

Twenty-two Approaches in Practice

APPROACH	REPLACING											REWEIGHTING						REWEIGHTING AND REPLACING (or vice versa)				
	Within-survey		Survey and External Data									Within-survey		Survey and External Data				Within-survey	Survey and External Data			
METHOD	Semiparametric	Nonparametric Imputation	Semiparametric				Nonparametric					Weighting Class Adjustment	Model Weight Adjustment	Poststratification	Reweighting Top w/Exogenous Threshold		Reweighting Top w/Endogenous Threshold	Model Weight Adjustment and Semiparametric	Reweighting Top w/Exogenous Threshold and Semiparametric	Reweighting Top w/ Endogenous Threshold and Rescaling	Nonparametric and Reweighting Top w/ Exogenous Threshold	
			External Income Microdata	Regression-based Prediction of Income	External Income Totals	Reweighting Top with External Income Totals	Rescaling Top Incomes w/External Income Microdata	Rescaling Top Incomes w/Income Totals	Statistical Matching	Replacing Top with External Data in Full	Reweighting Top with External Income Microdata				External Income Microdata	Income Totals					External Income Microdata	External Income Totals
Applications	Cowell and Flachaire (2007); Burkhauser et al. (2012); Alfons et al. (2013); Hlasny and Verme (2018a; 2018b; 2021); Burkhauser et al. (2010)	Hirsch and Schumacher (2004); Bollinger and Hirsch (2006)	Alvaredo (2011), Burkhauser et al. (2012); Alvaredo and Londoño (2013); Jenkins (2017)	Van der Weide, Lakner and Ianchovichina (2018)	Lakner and Milanovic (2016)	Bourguignon (2018)	Piketty, Yang and Zucman (2019)	Altimir (1987), Bourguignon (2018)	Bach, Corneo, and Steiner (2009), Bach, Beznozka and Steiner (2016)	Bollinger et al, (2019); Flachaire et al.(2023)	Medeiros et al. (2018)	Harris (1977), Atkinson and Micklewright (1983)	Mistiaen and Ravallion (2003); Korinek et al. (2006; 2007), Hlasny and Verme (2018a; 2018b; 2021)	Atkinson and Micklewright (1983), Campos-Vazquez and Lustig (2019)	Flachaire et al. (2023)	Bourguignon (2018)	Blanchet et al. (2022), Flachaire et al. (2023)	Hlasny and Verme (2018a; 2022)	Atkinson (2007), Anand and Segal (2015)	Blanchet et al. (2022)	Burkhauser et al. (2018)	Bourguignon (2018)
Type of Data	Survey		Survey, Tax and Social Security	Survey and House Prices (or other variables that predict incomes)	Survey and National Accounts	Survey and National Accounts	Survey and Tax	Survey and National Accounts	Survey and Tax			Survey and Nonresponse Rate by Gegraphic Area	Survey and Nonresponse Rate by Primary Sampling Unit or Geographic Area	Survey, Census, Tax and Social Security	Survey, Tax and Social Security Survey	Survey and National Accounts	Survey and Tax	Survey and Nonresponse Rate by Primary Sampling Unit or Geographic Area	Survey and Tax		Survey and National Accounts
Assumes Common Support	No	Yes	No									Yes						No				
Weight of the Upper Tail and the Rest Intact	Yes											No										
Weights within Rest of Distribution Intact	Yes											No		No (Mechanical Uniform Downweighting)				No	No (Mechanical Uniform Downweighting)			
Observations (incomes) within Upper Tail Intact	No											Yes						No				
Absolute Poverty Indicators Intact	Yes											No										
Generates Corrected	Distribution	Microdata	Distribution	Distribution	Distribution	Distribution	Distribution	Microdata**	Microdata	Microdata*	Microdata	Microdata						Distribution	Distribution	Microdata**	Microdata**	Distribution

Source: Table 3, Lustig and Vigorito (2025).

Data Reconciliation: Surveys and Tax Data

The approaches that combine surveys with external sources require three types of reconciliation:

- Income concept:
  - Reconciling the income variable in household surveys and the external source entails using the same (or most similar) income concept: tax data includes taxable income so in survey one needs to exclude informal income, for example
  - Converting consumption into income in consumption-based surveys

- Unit of analysis:
  - Households, individuals, adults, tax units? Typically, empirical exercises use the population aged 15 or 20 and over because the tax data covers the adult population only

- Income-sharing unit:
  - Individuals or married couples (in the countries where there is joint tax filing)

Sensitivity of Results

- Sensitivity to threshold selection
- Sensitivity to correction approach

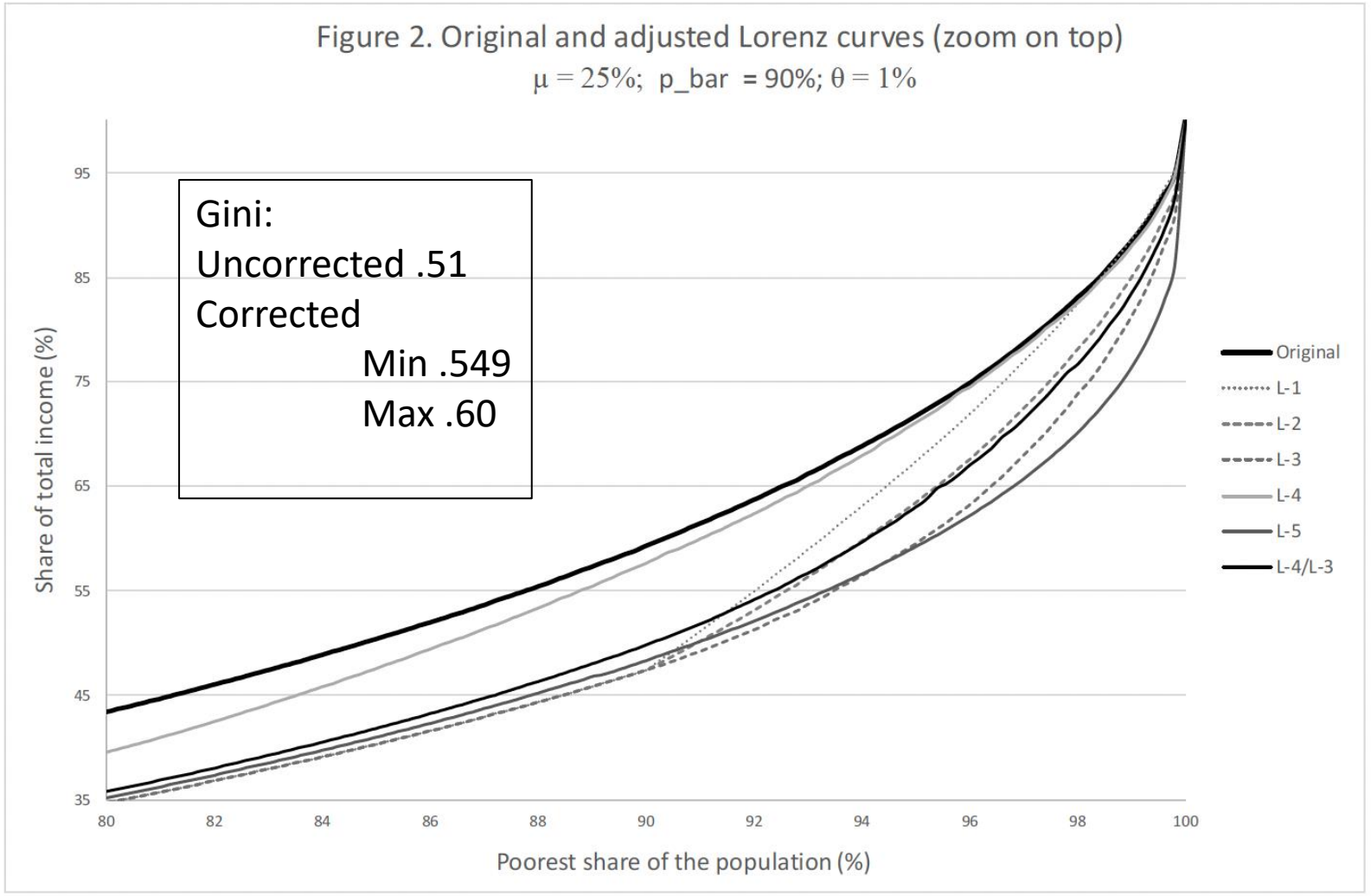
Gini Coefficient: Same Data, Same Correction Approaches, Different Thresholds

			Uncorrected	Threshold	REPLACING	REWEIGHTING	REWEIGHTING AND REPLACING
					Semiparametric	Model Weight Adjustment	Model Weight Adjustment and Semiparametric
Hlasny&Verme (2021)	US	2013	0.4725	1%	0.491	**0.5038**	0.483
				5%	**0.5792**	0.5038	0.5226

- How to select the threshold? Cowell & Flachaire (2015); Jenkins (2017)

- Sensitivity to threshold selection

- Sensitivity to correction approach

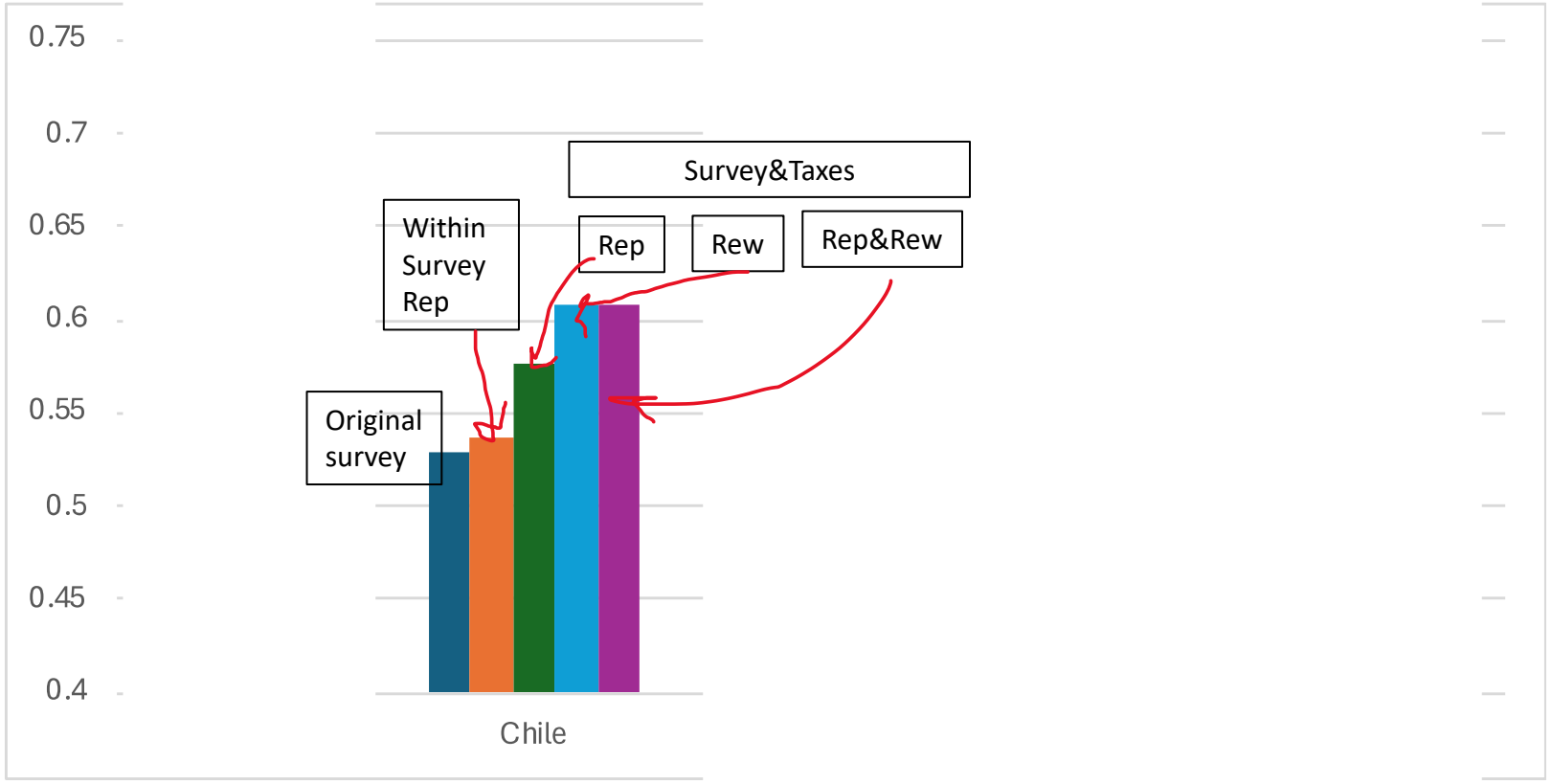


Mexico:

- Same data
  - Survey and National Accts
- Different correction Approaches
  - Rescaling
  - Reweighting
  - Reweighting within top
  - Rescaling and reweightin

Source: Bourguignon (2018)

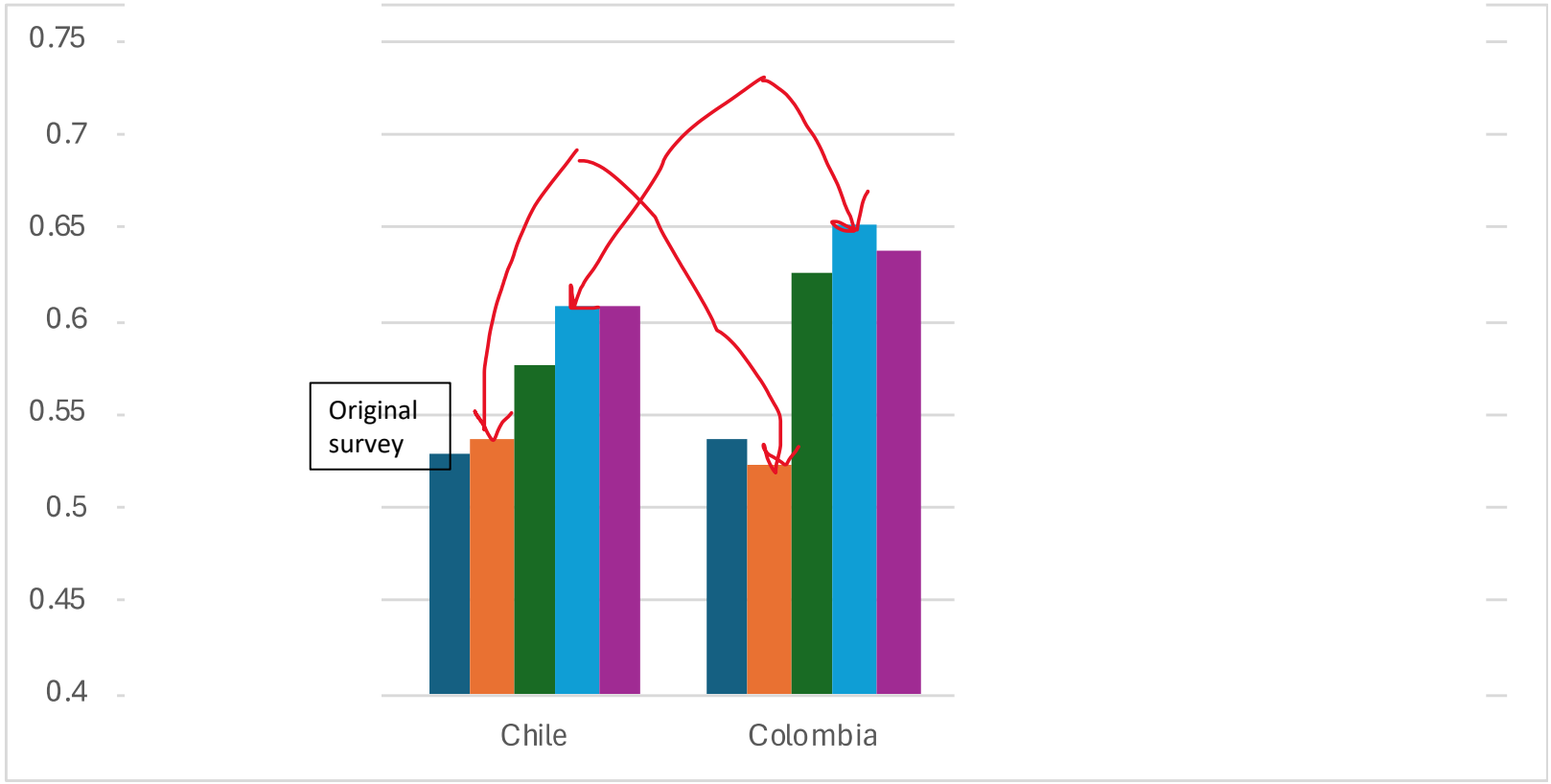
Gini Coefficient: Same Data, Different Correction Approaches



De Rosa, Lustig and Martinez Pabon (forthcoming)

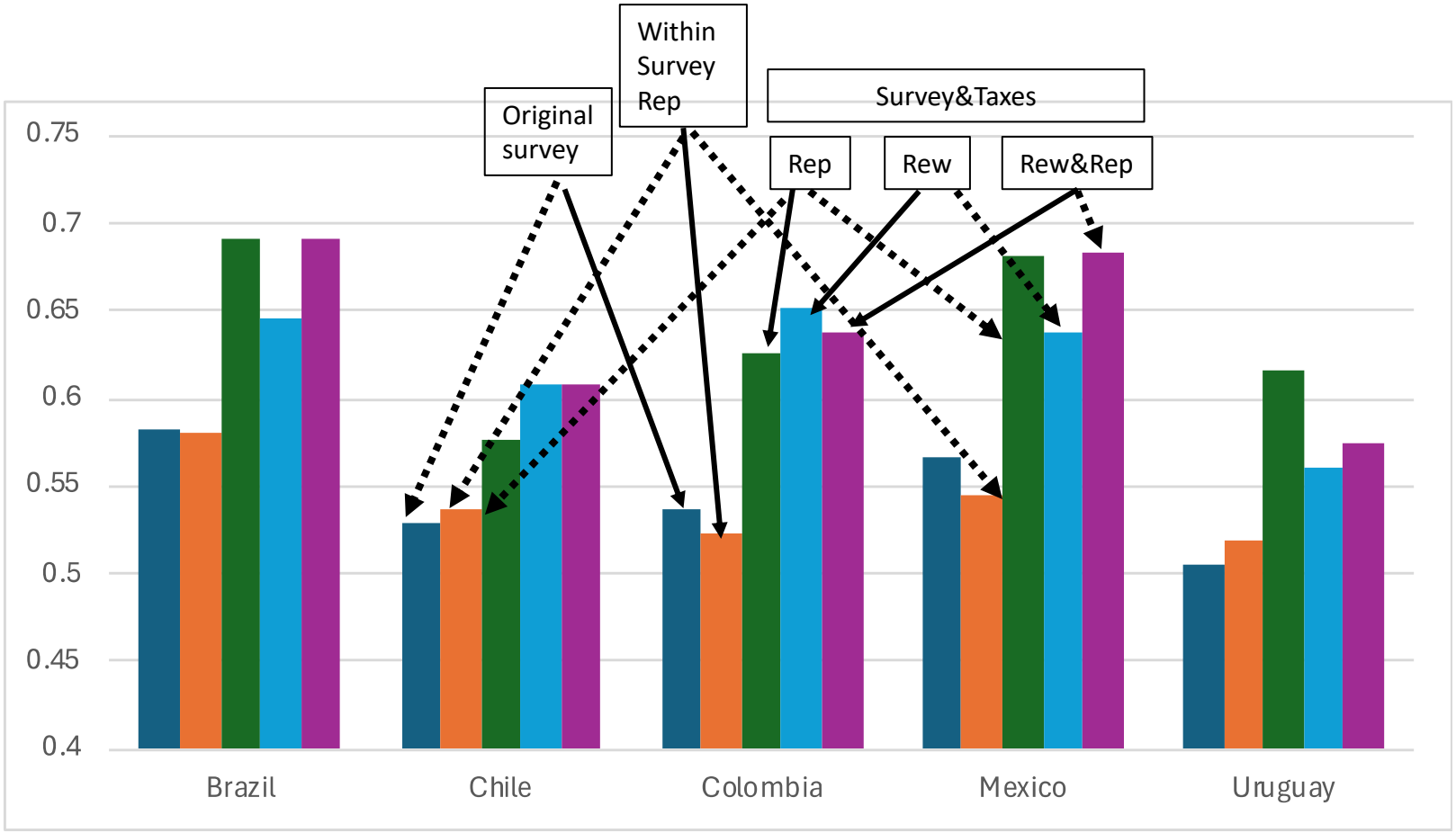
APPROACH	REPLACING											REWEIGHTING						REWEIGHTING AND REPLACING (or vice versa)				
	Within-survey		Survey and External Data									Within-survey		Survey and External Data				Within-survey	Survey and External Data			
METHOD	Semiparametric	Nonparametric Imputation	Semiparametric				Nonparametric					Weighting Class Adjustment	Model Weight Adjustment	Poststratification	Reweighting Top w/Exogenous Threshold		Reweighting Top w/Endogenous Threshold	Model Weight Adjustment and Semiparametric	Reweighting Top w/Exogenous Threshold and Semiparametric	Reweighting Top w/ Endogenous Threshold and Rescaling	Nonparametric and Reweighting Top w/ Exogenous Threshold	
			External Income Microdata	Regression-based Prediction of Income	External Income Totals	Reweighting Top with External Income Totals	Rescaling Top Incomes w/External Income Microdata	Rescaling Top Incomes w/Income Totals	Statistical Matching	Replacing Top with External Data in Full	Reweighting Top with External Income Microdata				External Income Microdata	Income Totals					External Income Microdata	External Income Totals
Applications	Cowell and Flachaire (2007); Burkhauser et al. (2012); Alfons et al. (2013); Hlasny and Verme (2018a; 2018b; 2021); Burkhauser et al. (2010)	Hirsch and Schumacher (2004); Bollinger and Hirsch (2006)	Alvaredo (2011), Burkhauser et al. (2012); Alvaredo and Londoño (2013); Jenkins (2017)	Van der Weide, Lakner and Ianchovichina (2018)	Lakner and Milanovic (2016)	Bourguignon (2018)	Piketty, Yang and Zucman (2019)	Altimir (1987), Bourguignon (2018)	Bach, Corneo, and Steiner (2009), Bach, Beznozka and Steiner (2016)	Bollinger et al, (2019); Flachaire et al.(2023)	Medeiros et al. (2018)	Harris (1977), Atkinson and Micklewright (1983)	Mistiaen and Ravallion (2003); Korinek et al. (2006; 2007), Hlasny and Verme (2018a; 2018b; 2021)	Atkinson and Micklewright (1983), Campos-Vazquez and Lustig (2019)	Flachaire et al. (2023)	Bourguignon (2018)	Blanchet et al. (2022), Flachaire et al. (2023)	Hlasny and Verme (2018a; 2022)	Atkinson (2007), Anand and Segal (2015)	Blanchet et al. (2022)	Burkhauser et al. (2018)	Bourguignon (2018)
Type of Data	Survey		Survey, Tax and Social Security	Survey and House Prices (or other variables that predict incomes)	Survey and National Accounts	Survey and National Accounts	Survey and Tax	Survey and National Accounts	Survey and Tax			Survey and Nonresponse Rate by Gegraphic Area	Survey and Nonresponse Rate by Primary Sampling Unit or Geographic Area	Survey, Census, Tax and Social Security	Survey, Tax and Social Security Survey	Survey and National Accounts	Survey and Tax	Survey and Nonresponse Rate by Primary Sampling Unit or Geographic Area	Survey and Tax		Survey and National Accounts
Assumes Common Support	No	Yes	No									Yes						No				
Weight of the Upper Tail and the Rest Intact	Yes											No										
Weights within Rest of Distribution Intact	Yes											No		No (Mechanical Uniform Downweighting)				No	No (Mechanical Uniform Downweighting)			
Observations (incomes) within Upper Tail Intact	No											Yes						No				
Absolute Poverty Indicators Intact	Yes											No										
Generates Corrected	Distribution	Microdata	Distribution	Distribution	Distribution	Distribution	Distribution	Microdata**	Microdata	Microdata*	Microdata	Microdata						Distribution	Distribution	Microdata**	Microdata**	Distribution

Gini Coefficient: Same Data, Different Correction Approaches, No Systematic Pattern



De Rosa, Lustig and Martinez Pabon (forthcoming)

Gini Coefficient: Same Data, Different Correction Approaches, No Systematic Pattern



De Rosa, Lustig and Martinez Pabon (forthcoming)

Same Data, Different Correction Approaches: Sensitivity of Results

Author	Country	Year	External Data	Uncorrected	Threshold	REPLACING					REWEIGHTING			REWEIGHTING AND REPLACING (or vice versa)		
						Within-survey	Survey and External Data				Within-survey	Survey and External Data		Within-survey	Survey and External Data	
						Semiparametric	Semiparametric	Semiparametric	Nonparametric		Model Weight Adjustment	Exogenous Threshold	Endogenous Threshold	Model Weight Adjustment and Semiparametric	Endogenous Threshold and Rescaling	Nonparametric and Exogenous Threshold
								Reweighting Top with External Data	Rescaling Top Incomes	Replacing Top with External Data in Full						
Hlasny and Verme	EU (2011)		none	38.23	1%		-	-	-	-	44.31	-	-	44.31	-	-
Hlasny&Verme (2021)	US	2013	none	0.4725	1%	0.491	-	-	-	-	**0.5038**	-	-	0.483	-	-
		2013			5%	**0.5792**	-	-	-	-	0.5038	-		0.5226	-	-
	Egypt (2009)		none	35.56	1%		-	-	-	-	41.16	-	-	41.15	-	-
Bourguignon (2018)	Mexico	2009	Nat Acc	0.51	1%	-	-	0.599	**0.6**	-	-	0.549	-	-	-	0.587
Flachaire et al. (2023)	Uruguay	2012	Tax	0.382	10%&72%	-	-		0.44	0.44	-	**0.442**	-	-	0.435	-
De Rosa et al (forth.)	Brazil		Tax	0.582	1%	0.581	**0.692**	-	-	-	-	-	0.646	-	0.691	-
	Chile		Tax	0.529		0.537	0.576	-	-	-	-	-	0.609	-	**0.609**	-
	Colombia		Tax	0.538		0.523	0.627	-	-	-	-	-	**0.652**	-	0.639	-
	Mexico		Tax	0.567		0.545	0.681	-	-	-	-	-	0.638	-	**0.684**	-
	Uruguay		Tax	0.505		0.519	**0.617**	-	-	-	-	-	0.561	-	0.575	-

In grey highlight, corrected lower than uncorrected Gini

Source: Table 4, Lustig and Vigorito (2025).

Choosing Correction Approach(es)

Analytical Decision Tree

1) To reweight or not to reweight?
   - If yes, model-based recalibration or poststratification?

2) To replace or not to replace?
   - If yes, semiparametric or nonparametric?

3) To combine survey with external data or not to combine?
   - If yes,
     - Tax data or National Accounts?
     - Semiparametric or nonparametric?

Analytical Decision Tree

1) To reweight or not to reweight?

➢ Check need with Korinek et al. test and/or the Groves test

- If answer is yes, recalibrate weights using (we recommend) Korinek, Mistiaen and Ravallion (2006; 2007)

✓ Advantages: it keeps statistical integrity of survey; covariates; statistical significance tests can be applied

✓Limitations: if support is not common (i.e., not a single rich individual made it into the sample), correction will be limited

✓ Challenges: requires nonresponse rates by PSU; computationally complex; perhaps hard to apply "bulk" for a time series for many countries simultaneously

1) To reweight or not to reweight?

- Korinek et al test:
  - Obtain the rate of unit nonresponse by Primary Sampling Unit (PSU) or the more disaggregated geographic unit possible within the sampling frame.
  - Calculate the average income for each geographic unit based on the observations in the survey
  - Plot or regress the rate of unit nonresponse against the average income
  - If the plot is not a horizontal line, then unit nonresponse is not missing at random, and reweighting is a necessary correction step to reduce bias
  - Evidence of underrepresentation of the rich would be indicated if the plot is upward sloping or U-shaped
- Groves test:
  - Compare the respondent-based distribution of the variable of interest in the survey with the distribution from another more accurate source (Groves, 2006, p. 655)
    - Example: proportion of individuals above the income threshold corresponding to the top 1 percent in UK survey is very similar to proportion above that same income level in the tax data (Burkhauser et al. , 2018)

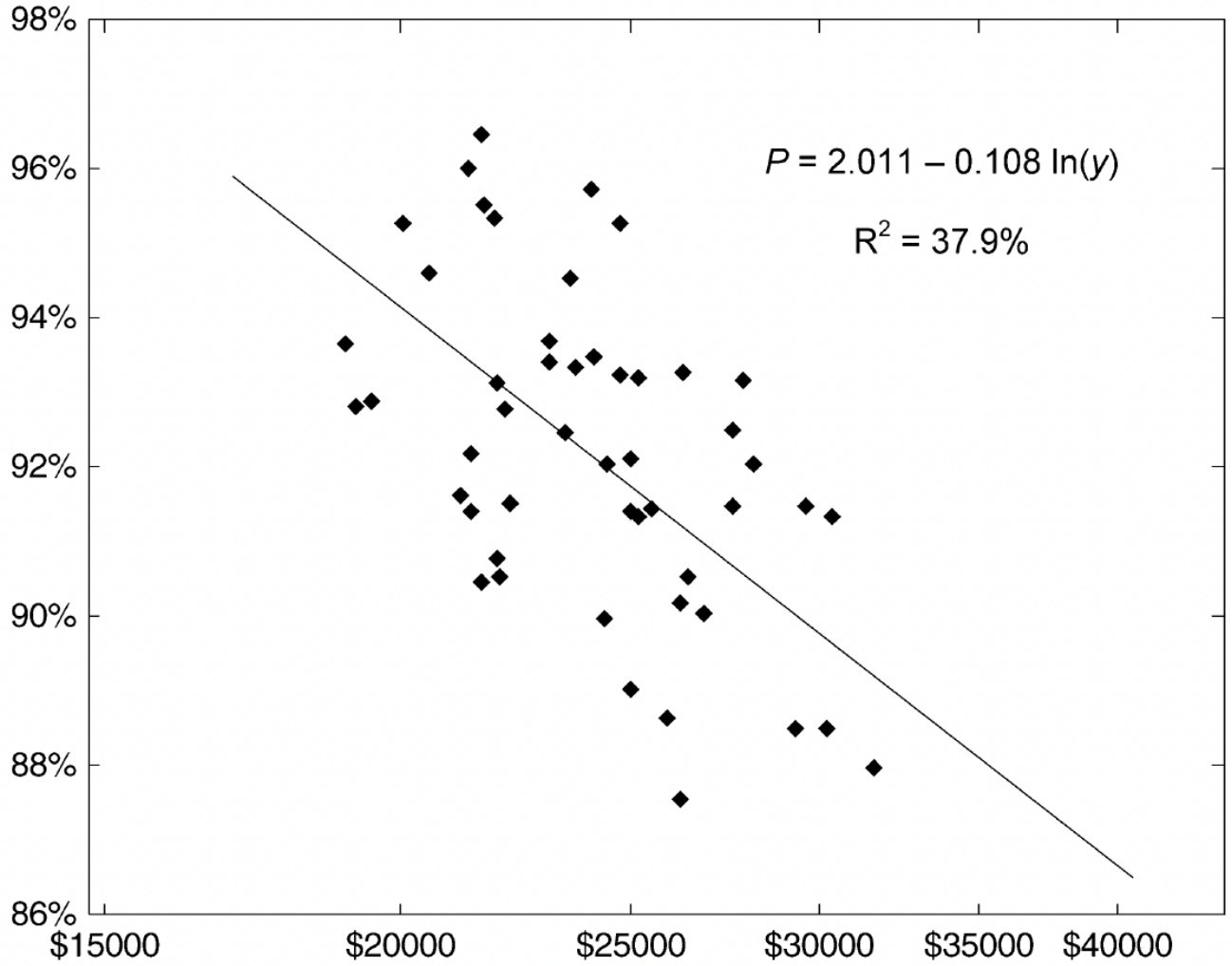


Figure 2. Probability of response against per-capita income by state.

Source: Korinek et al. (2006)

2) To replace or not to replace?

- Check presence of typical upper tail issues : sparseness, data contamination, high-leverage observations, underreporting, top coding, etc. (Cowell & Flachaire, 2015)

  - If answer is yes, apply within-survey semiparametric replacement methods after recalibrating weights, if necessary (previous slide); examples: Burkhauser et al. (2012); Hlasny and Verme (2018a, 2018b)

    ✓Advantages: can obtain corrected distribution for the same income concept, unity of analysis, and income-sharing unit as in household survey

    ✓Limitations:

    - Of all semiparametric methods: loss of covariates (perhaps it can be overcome)
    - Within-survey estimates of parameters may yield limited correction

3) To combine survey with external data or not to combine?

- ➢ Assess if support of survey and target population differ by inspection or other methods
  - If answer is yes, which data? what method?
    - If there is reliable tax data, that could be first choice
    - Method:
      - Semiparametric: example, Jenkins (2017)
        - ✓ Advantages: solid strategy in selecting the threshold, parametric model, and estimation method of the parametric model; robustness checks and sensitivity analysis
        - ✓ Limitations:
          - Of all semiparametric methods: loss of covariates
          - Cannot obtain (in general) corrected distribution for the same income concept, unity of analysis, and income-sharing unit as in household survey
      - Nonparametric: examples, statistical matching (Bach et al. (2009) and with linked data, "hybrid" distributions (Bollinger et al., 2019; Flachaire et al. 2023)
        - ✓ Advantages:
          - Can obtain corrected distribution for the same income concept, unity of analysis, and income-sharing unit as in household survey
          - Covariates can be preserved
        - ✓ Limitations:
          - Matching is based on observable variables while nonresponse and misreporting may be influenced by unobservables
          - Linked data may have linkage errors; which of the linked data reports the true income?

3) To combine survey with external data or not to combine?

- If only National Accounts income totals are available, more reliable or want to use them as control totals
- Method:
  - Semiparametric and Nonparametric
    - Example: Bourguignon (2018); obtains a range of corrected estimates given that there is no way to determine which of the alternatives is closer to the true distribution

✓Advantages:

- If nonparametric, as all rescaling methods:
  - Can obtain corrected distribution for the same income concept, unity of analysis, and income-sharing unit as in household survey
  - Covariates can be preserved

✓Limitations:

- Allocation of gap relies entirely on assumptions of how that gap is distributed since you can only compare totals

Conclusions

Conclusions

- Household surveys often fail to accurately capture the incomes of the richest individuals, leading to biased and imprecise inequality measures
- We identified 22 distinct correction approaches
- Inequality measures can vary significantly after correction, both in levels and trends, with no consistent pattern across different correction approaches
- Unfortunately, there are no statistical tests or general calibration mechanisms to rank them

➢Warning against a mechanical application of methods

Conclusions

- Provided broad guidelines on approach selection process

➢ Given the limitations of all correction methods, systematic robustness checks and reporting a range (or bands) of corrected inequality measures rather than single point estimates

- For example:
  - Comparing within-survey approaches with combined survey-external data methods
  - Analyzing sensitivity to:
    - Threshold selection
    - External data types
    - Estimation methodologies

Conclusions

- The future:
  - Methodological innovations for testing and calibrating results
  - Linked data
  - Machine learning?

References

- Abowd, J. M., and Stinson, M. H. (2013). Estimating measurement error in annual job earnings: A comparison of survey and administrative data. Review of Economics and Statistics, 95(5), 1451-1467.
- Alfons, A., Templ, M. and Filzmoser, P. (2013). Robust estimation of economic indicators from survey samples based on Pareto tail modelling. Journal of the Royal Statistical Society 62 (C), pp. 271–86.
- Altimir, O. (1979). La dimensión de la pobreza en América Latina. Cuadernos de la CEPAL N27, Santiago de Chile.
- Altimir, O. (1987). Income Distribution Statistics in Latin America and their Reliability. Review of Income and Wealth, Vo. 33, Issue 2, June, pp. 111-155.
- Alvaredo, F. (2007). The rich in Argentina over the twentieth century: From the Conservative Republic to the Peronist experience and beyond 1932-2004.
- Alvaredo, F. (2011). A Note on the Relationship Between Top Income Shares and the Gini Coefficient," Economics Letters 110 (3), pp. 274-277.
- Alvaredo, F. and Londoño-Velez, J. (2013). High Incomes and Personal Taxation in a Developing Economy: Colombia 1993-2010. CEQ Working Paper 12, Center for Inter-American Policy and Research and Department of Economics, Tulane University and Inter-American Dialogue.
- Alvaredo, F., Assouad, L., and Piketty, T. (2019). Measuring lnequality in the Middle East 1990–2016: The World's Most Unequal Region?. Review of Income and Wealth, 65(4), 685-711.

- Anand, S., and Segal, P. (2015). The global distribution of income. In A. B. Atkinson, and F. Bourguignon (Eds.). Handbook of income distribution (2A, pp. 937–979). Amsterdam: North-Holland.
- Atkinson, A. B. (2007). Measuring Top Incomes: Methodological Issues, in Atkinson, A. B. and T. Piketty (eds.), Top Incomes over the Twentieth Century - A Contrast between Continental European and English-Speaking Countries, Oxford and New York: Oxford University Press.
- Atkinson, A. B. and Micklewright, J. (1983). On the Reliability of Income Data in the Family Expenditure Survey 1970- 1977. Journal of the Royal Statistical Society. Series A (General) 146, no. 1 (1983): 33-61. doi:10.2307/2981487.
- Atkinson, A. B. and Piketty, T. (2007). Top Incomes in the Twentieth Century, Oxford: Oxford University Press.
- Atkinson, A. B. and Piketty, T. (2010). Top Incomes. A Global Perspective, Oxford: Oxford University Press.
- Atkinson, A. B., Piketty, T., and Saez, E. (2011). Top Incomes in the Long Run of History. Journal of Economic Literature 49 (1), pp. 3-71.
- Bach, S., Beznoska, M., and Steiner, V. (2016). Who bears the tax burden in Germany? Tax structure slightly progressive. DIW Economic Bulletin, 6(51/52), 601-608.
- Bach, S., Corneo, G., and Steiner, V. (2009). From Bottom To Top: The Entire Income Distribution In Germany, 1992-2003. Review of Income and Wealth 55(2), pp. 303-330, 06.

- Biemer, P. and Christ, S. (2008). Weighting Survey Data, Chapter 17, in De Leeuw, E., J. Hox, and D. Dillman International Handbook of Survey Methodology. Great Britain: Psychology Press.
- Blanchet T., Chancel L., Flores I. and Morgan M. (2024) Distributional National Accounts Guidelines. World Inequality Lab
- Blanchet, T., Flores, I., and Morgan, M. (2022). The Weight of the Rich: improving surveys using tax data. The Journal of Economic Inequality, 20(1), 119-150.
- Bollinger, C. R., and Hirsch, B. T. (2006). Match bias from earnings imputation in the Current Population Survey: The case of imperfect matching. Journal of Labor Economics, 24(3), 483-519.
- Bollinger, C. R., Hirsch, B. T., Hokayem, C. M., and Ziliak, J. P. (2019). Trouble in the tails? What we know about earnings nonresponse 30 years after Lillard, Smith, and Welch. Journal of Political Economy, 127(5), 2143-2185.
- Bourguignon, F. (2018). Simple adjustments of observed distributions for missing income and missing people. The Journal of Economic Inequality, 16, 171-188.
- Burkhauser, R. V., Feng, S., Jenkins, S. P., and Larrimore, J. (2012). Recent trends in top income shares in the USA: reconciling estimates from March CPS and IRS tax return data. Review of Economics and Statistics 94, pp. 371–88.

- Burkhauser, R. V., Hérault, N., Jenkins, S. P., and Wilkins, R. (2018). Survey Under-Coverage of Top Incomes and Estimation of Inequality: What is the Role of the UK's SPI Adjustment?. Fiscal Studies, 39(2), 213-240.
- Campos-Vazquez, R. M., and Lustig, N. (2019). Labour income inequality in Mexico: Puzzles solved and unsolved. Journal of Economic and Social Measurement, 44(4), 203-219.
- Chancel, L., Cogneau, D., Gethin, A., Myczkowski, A., & Robilliard, A. S. (2023). Income inequality in Africa, 1990–2019: Measurement, patterns, determinants. World Development, 163, 106162.
- Cowell, F. A. and Flachaire, E. (2007). Income Distribution and Inequality measurement: The Problem of Extreme Values. Journal of Econometrics 141(2), pp. 1044–1072.
- Cowell, F. A. and Flachaire, E. (2015). Statistical Methods for Distributional Analysis, Chapter 6 in Atkinson, A. B. and F. Bourguignon (eds.), Handbook of Income Distribution, Vol. 2, North-Holland, Amsterdam: Elsevier.
- De Rosa, M., Lustig, N., and Martinez Pabon, V. (forthcoming). Fiscal redistribution when accounting for the rich in Latin America. CEQ Working Paper Series.
- De Rosa, M., Flores, I., and Morgan, M. (2024). More unequal or not as rich? Revisiting the Latin American exception. World Development, 184, 106737.
- Deaton, A. (2005). Measuring Poverty in a Growing World (or Measuring Growth in a Poor World). The Review of Economics and Statistics 87 (1), pp. 1-19.

- Department for Work and Pensions, UK. (2015). Households Below Average Income An Analysis of the Income Distribution 1994/95–2013/14. London: Department for Work and Pensions.
- Flachaire, E., Lustig, N., and Vigorito, A. (2023). Underreporting of top incomes and inequality: a comparison of correction methods using simulations and linked survey and tax data. Review of Income and Wealth, 69(4), 1033-1059.
- Groves, R. M. and Couper, M. P. (1998). Nonresponse in Household Interview Surveys, New York: Wiley.
- Groves, R. M., Fowler, F. J., Couper, M. P., Lepkowski, J. M., Singer, E., Tourangeau, R. (2009). Survey Methodology, New Jersey: John Wiley & Sons.
- Harris R.P. (1977). Differential response in the Family Expenditure Survey: the effect of estimates of the redistribution of income, Statistical News 39.
- Hirsch, B. T., and Schumacher, E. J. (2004). Match bias in wage gap estimates due to earnings imputation. Journal of Labor Economics, 22(3), 689-722.
- Hlasny, V. (2021). Parametric representation of the top of income distributions: Options, historical evidence, and model selection. Journal of Economic Surveys, 35(4), 1217-1256.
- Hlasny, V. and P. Verme (2018a). Top Incomes and Inequality Measurement: A Comparative Analysis of Correction Methods Using the EU SILC Data. Econometrics 6(2):30.

- Hlasny, V., and Verme, P. (2018b). Top incomes and the measurement of inequality in Egypt. The World Bank Economic Review, 32(2), 428-455.
- Hlasny, V., and Verme, P. (2021). The Impact of Top Incomes Biases on the Measurement of Inequality in the United States. Oxford Bulletin of Economics and Statistics, 84(4), 749-788.
- Hümbelin, O., & Farys, R. (2016). The Suitability of Tax Data to Study Trends in Inequality—A theoretical and empirical review with tax data from Switzerland. Research in Social Stratification and Mobility, 44, 136–150.
- Hyslop, D. R., and Townsend, W. (2020). Earnings dynamics and measurement error in matched survey and administrative data. Journal of Business & Economic Statistics, 38(2), 457-469.
- Jenkins, S. P. (2017). Pareto Models, Top Incomes and Recent Trends in UK Income Inequality. Economica 84 (334), pp. 261-289.
- Jenkins, S. P., and Rios-Avila, F. (2020). Modelling errors in survey and administrative data on employment earnings: Sensitivity to the fraction assumed to have error-free earnings. Economics Letters, 192, 109253.
- Kapteyn, A., and Ypma, J. Y. (2007). Measurement error and misclassification: A comparison of survey and administrative data. Journal of Labor Economics, 25(3), 513-551.
- Kim (2014). Income inequality in Korea. Piketty Fever and Income Distribution in Korea: Reality and Prescription. EAF Policy Debates 13, East Asia Foundation.

- Korinek, A., Mistiaen, J. A., and Ravallion, M. (2007). An Econometric Method of Correcting for UnitNonresponse Bias in Surveys. Journal of Econometrics,136(1), 213–235.
- Korinek, A., Mistiaen, J.A. and Ravallion, M. (2006). Survey nonresponse and the distribution of income, Journal of Economic Inequality, 4, 33-55.
- Kuznets, S. (1953). Economic Change, New York: Norton.
- Lakner, C. and Milanovic, B. (2016). Global Income Distribution: From the Fall of the Berlin Wall to the Great Recession. The World Bank Economic Review, Volume 30, Issue 2, Pages 203–232.
- Lustig N. (2019). The “Missing Rich” in Household Surveys: Causes and Correction Approaches, Working Paper 75, Commitment to Equity, Tulane University.
- Including the Rich in Income Inequality Measures: An Assessment of Correction Approaches. Working Paper 2503. Tulane University. February. https://ideas.repec.org/p/tul/wpaper/2503.html
- Medeiros, M., de Castro Galvão, J., and de Azevedo Nazareno, L. (2018). Correcting the underestimation of top incomes: combining data from income tax reports and the Brazilian 2010 census. Social Indicators Research, 135, 233-244.
- Medeiros, M., & De Souza, P. H. F. (2015). The rich, the affluent and the top incomes. Current Sociology, 63(6), 869–895

- Meijer, E., Rohwedder, S., and Wansbeek, T. (2012). Measurement error in earnings data: using a mixture model approach to combine survey and register data. Journal of Business & Economic Statistics, 30(2), 191-201.
- Mistiaen, J. A., and Ravallion, M. (2003). Survey compliance and the distribution of income (Vol. 2956). World Bank Publications.
- Piketty, T., Saez, E., and Zucman, G. (2022). Twenty years and counting: Thoughts about measuring the upper tail. The Journal of Economic Inequality, 20(1), 255-264.
- Piketty, T., Yang, L., and Zucman, G. (2019). Capital Accumulation, Private Property, and Rising Inequality in China, 1978–2015. American Economic Review, 109 (7): 2469-96.
- Ravallion, M. (2022). Missing top income recipients. The Journal of Economic Inequality, 20(1), 205-222.
- Rubin, D. (1987). Multiple Imputation for Nonresponse in Surveys, New York: Wiley.
- Szekely, M., and Hilgert, M. (2007). What's behind the inequality we measure? An investigation using Latin American data. Oxford Development Studies, 35(2), 197-217.
- Van der Weide, R., Lakner, C., and Ianchovichina, E. (2018). Is Inequality Underestimated in Egypt? Evidence from House Prices. The Review of Income and Wealth, Vol. 64, Issue s1, pp. S55-S79.
- Yonzan, N., Milanovic, B., Morelli, S., and Gornick, J. (2022). Drawing a line: Comparing the estimation of top incomes between tax data and household survey data. The Journal of Economic Inequality, 20(1), 67-95.
- Zizzamia, R., David, A., and Leibbrandt, M. (2021). Inequality in sub-Saharan Africa: A review paper. AFD Research Papers, (207), 1-33.

Thank you!