When Measure Matters: Coresidence Bias and Intergenerational Mobility Revisited

Ercio A. Munoz
CUNY Graduate Center and
The World Bank

Mariel Siravegna
Georgetown University

December 2021

Background and Motivation

What is Intergenerational Mobility (IGM)

- IGM studies the link between the socioeconomic conditions of the individuals and those of their parents.

- "Absolute" mobility measures progress in absolute terms with respect to parents. It matters as a measure of long-term improvement of living standards for all.
  - Share of children with more years of schooling than parents
  - Bottom upward mobility-primary: $Pr(C^c \geq primary | C^p < primary)$
  - Bottom upward mobility-secondary: $Pr(C^c \geq secondary | C^p < secondary)$

- "Relative" mobility measures progress in relative position with respect to peers compared to the position of parents relative to their peers. It matters for economic growth, and both can reinforce each other.
  - IGRC= OLS estimate of the slope (β) in $S^y = \alpha + \beta S^o$
  - IGPC= Pearson correlation coefficient (ρ), where $\rho = Corr(S^o, S^y)$
  - BHQ4= Prob. of reaching top quartile if parents are in bottom half, $Pr(R^y > 75 | R^o \leq 50)$

What is Intergenerational Mobility (IGM)

- IGM studies the link between the socioeconomic conditions of the individuals and those of their parents.

- "Absolute" mobility measures progress in absolute terms with respect to parents. It matters as a measure of long-term improvement of living standards for all.
  - Share of children with more years of schooling than parents
  - Bottom upward mobility-primary: $Pr(C^c \geq primary | C^p < primary)$
  - Bottom upward mobility-secondary: $Pr(C^c \geq secondary | C^p < secondary)$

- "Relative" mobility measures progress in relative position with respect to peers compared to the position of parents relative to their peers. It matters for economic growth, and both can reinforce each other.
  - IGRC= OLS estimate of the slope (β) in $S^y = \alpha + \beta S^o$
  - IGPC= Pearson correlation coefficient (ρ), where $\rho = Corr(S^o, S^y)$
  - BHQ4= Prob. of reaching top quartile if parents are in bottom half, $Pr(R^y > 75 | R^o \leq 50)$

What is Intergenerational Mobility (IGM)

- IGM studies the link between the socioeconomic conditions of the individuals and those of their parents.

- "Absolute" mobility measures progress in absolute terms with respect to parents. It matters as a measure of long-term improvement of living standards for all.
  - Share of children with more years of schooling than parents
  - Bottom upward mobility-primary: $Pr(C^c \geq primary | C^p < primary)$
  - Bottom upward mobility-secondary: $Pr(C^c \geq secondary | C^p < secondary)$

- "Relative" mobility measures progress in relative position with respect to peers compared to the position of parents relative to their peers. It matters for economic growth, and both can reinforce each other.
  - IGRC= OLS estimate of the slope (β) in $S^y = \alpha + \beta S^o$
  - IGPC= Pearson correlation coefficient (ρ), where $\rho = Corr(S^o, S^y)$
  - BHQ4= Prob. of reaching top quartile if parents are in bottom half, $Pr(R^y > 75 | R^o \leq 50)$

Measurement problems

- Measurement of IGM in education requires information about children's and parents' educational attainment. This information can be obtained asking the respondents about their parents (or abour their children) or through panel data (either a survey or administrative data)

- However, many times the only data available are coresident samples
  - Samples with this link only available for individuals living with their parents

- Since the decision to live with parents is not random, the use of coresident samples may yield biased estimates of Intergenerational Mobility
  - This issue is the focus of this paper

Measurement problems

- Measurement of IGM in education requires information about children's and parents' educational attainment. This information can be obtained asking the respondents about their parents (or abour their children) or through panel data (either a survey or administrative data)

- However, many times the only data available are coresident samples
  - Samples with this link only available for individuals living with their parents

- Since the decision to live with parents is not random, the use of coresident samples may yield biased estimates of Intergenerational Mobility
  - This issue is the focus of this paper

Measurement problems

- Measurement of IGM in education requires information about children's and parents' educational attainment. This information can be obtained asking the respondents about their parents (or abour their children) or through panel data (either a survey or administrative data)
- However, many times the only data available are coresident samples
  - Samples with this link only available for individuals living with their parents
- Since the decision to live with parents is not random, the use of coresident samples may yield biased estimates of Intergenerational Mobility
  - This issue is the focus of this paper

Recent work on IGM relevant to this paper

First, many recent papers use coresident samples to measure IGM:

Article	Coverage	Data and Sample	Indicators
Alesina et al. 2021	Africa	69 censuses (aged 14-25)	BUM, TDM
Alesina et al. 2020	Africa	37 censuses and 1 hh. survey (aged 14-18)	BUM, TDM
Asher et al. 2021	India	2011-12 SECC Census (aged 20-23)	BUM, TDM (interval)
Card et al. 2018	US	Census 1940 (aged 14-18 and 14-16)	BUM
Derenoncourt 2021	US	Census 1940 (aged 14-18)	BUM
Dodin et al. 2021	Germany	Microcensuses 1997-2018 (aged 17-21)	IGIG, Q5/Q1, Q1
Feigenbaum 2018	Iowa	Census 1915 Iowa and 1940 US (aged 3-17)	IGRC
Geng 2020	China	Census 1982, 1990, and 2000 (aged 23-32)	IGRC, IGPC, IGSC
Hilger 2016	US	Censuses from 1940 to 2000 (aged 26-29)	IGRC, IGRI
Van der Weide et al. 2021	153 countries	Household surveys (aged 21-25)	YOS, CAT, IGRC, IGPC
Van der Weide et al. 2020	Brazil	Census 2010 (aged 20-24)	IGRC, IGPC, YOS, IGRI

Recent work on IGM relevant to this paper

However, the literature about the impact of coresidence bias is scarce. An important exception is:

- **When Measure Matters: Coresidency, Truncation Bias, and Intergenerational Mobility in Developing Countries** Emran, M. S., Greene, W., Shilpi, F. (2018) *Journal of Human Resources*
  - *"The evidence and analysis in this paper thus provide a strong rationale for focusing on IGC as a measure of intergenerational mobility in the context of developing countries"*
  - *"Our analysis also suggest that the IGC estimates are much less sensitive to the variation in coresidency rates compared to the IGRC estimates"*
  - *"Much progress could made with the imperfect data if researchers move away from the current emphasis on IGRC and use IGC as the appropriate measure instead"*

This conclusion is relevant for the empirical literature

Two recent surveys about IGM in developing countries put them in this way:

- Torche (2019): *"If older co-resident children are included in the analysis, this induces the risk of bias insofar as children who continue to live with parents after late adolescence might not be a representative sample of their cohort. Emran et al. (2018) show that* ***co-residence bias affects IER much more strongly than it does IEC****. Selection bias induced by selecting co-resident children beyond their late adolescence is a concern even if the sample is restricted to children who are young adults"*

- Emran and Shilpi (2021): *"The evidence thus suggests that, for researchers working with the surveys readily available in developing countries,* ***it is better to rely on IGC as the measure of relative mobility****, and the current reliance on IGRC as the preferred measure seems ill-advised"*

This conclusion is relevant for the empirical literature

Two recent surveys about IGM in developing countries put them in this way:

- Torche (2019): *"If older co-resident children are included in the analysis, this induces the risk of bias insofar as children who continue to live with parents after late adolescence might not be a representative sample of their cohort. Emran et al. (2018) show that **co-residence bias affects IER much more strongly than it does IEC**. Selection bias induced by selecting co-resident children beyond their late adolescence is a concern even if the sample is restricted to children who are young adults"*

- Emran and Shilpi (2021): *"The evidence thus suggests that, for researchers working with the surveys readily available in developing countries, **it is better to rely on IGC as the measure of relative mobility**, and the current reliance on IGRC as the preferred measure seems ill-advised"*

This paper addresses the following questions

- Is the intergenerational correlation coefficient (IGPC) a less biased measure than the intergenerational regression coefficient (IGRC)? As a recent influential paper concludes
- How does coresidence bias affect different measures of intergenerational mobility?
- Can we use a set of estimates computed with coresident samples to rank economies by IGM across time and space?

This paper addresses the following questions

- Is the intergenerational correlation coefficient (IGPC) a less biased measure than the intergenerational regression coefficient (IGRC)? As a recent influential paper concludes
- How does coresidence bias affect different measures of intergenerational mobility?
- Can we use a set of estimates computed with coresident samples to rank economies by IGM across time and space?

This paper addresses the following questions

- Is the intergenerational correlation coefficient (IGPC) a less biased measure than the intergenerational regression coefficient (IGRC)? As a recent influential paper concludes
- How does coresidence bias affect different measures of intergenerational mobility?
- Can we use a set of estimates computed with coresident samples to rank economies by IGM across time and space?

Preview of findings

- We re-examine Emran et al. (2018)'s conclusion and offer empirical evidence against it.
  - IGPC is not always more robust than IGRC

- In addition, we show that even if the bias between both indicators is small and similar for each birth cohort separately, when we pool all of them the bias moves dramatically in favor of IGPC

- We provide novel evidence of the size of the coresidence bias for a large set of IGM indicators used in recent literature:
  - We find varying levels of coresidence bias going from less 1% to more than 10%
  - Indicators of absolute mobility recently used in the literature show low bias and produce reliable rankings
  - However, some indicators (relative) with minimal bias produce high levels of re-ranking that make them uninformative to rank economies (e.g., CER050)
  - In contrast, other indicators with large bias produce more reliable rankings (e.g., IGRC)

Preview of findings

- We re-examine Emran et al. (2018)'s conclusion and offer empirical evidence against it.
  - IGPC is not always more robust than IGRC
- In addition, we show that even if the bias between both indicators is small and similar for each birth cohort separately, when we pool all of them the bias moves dramatically in favor of IGPC
- We provide novel evidence of the size of the coresidence bias for a large set of IGM indicators used in recent literature:
  - We find varying levels of coresidence bias going from less 1% to more than 10%
  - Indicators of absolute mobility recently used in the literature show low bias and produce reliable rankings
  - However, some indicators (relative) with minimal bias produce high levels of re-ranking that make them uninformative to rank economies (e.g., CER050)
  - In contrast, other indicators with large bias produce more reliable rankings (e.g., IGRC)

Preview of findings

- We re-examine Emran et al. (2018)'s conclusion and offer empirical evidence against it.
  - IGPC is not always more robust than IGRC
- In addition, we show that even if the bias between both indicators is small and similar for each birth cohort separately, when we pool all of them the bias moves dramatically in favor of IGPC
- We provide novel evidence of the size of the coresidence bias for a large set of IGM indicators used in recent literature:
  - We find varying levels of coresidence bias going from less 1% to more than 10%
  - Indicators of absolute mobility recently used in the literature show low bias and produce reliable rankings
  - However, some indicators (relative) with minimal bias produce high levels of re-ranking that make them uninformative to rank economies (e.g., CER050)
  - In contrast, other indicators with large bias produce more reliable rankings (e.g., IGRC)

Preview of findings

- We re-examine Emran et al. (2018)'s conclusion and offer empirical evidence against it.
  - IGPC is not always more robust than IGRC
- In addition, we show that even if the bias between both indicators is small and similar for each birth cohort separately, when we pool all of them the bias moves dramatically in favor of IGPC
- We provide novel evidence of the size of the coresidence bias for a large set of IGM indicators used in recent literature:
  - We find varying levels of coresidence bias going from less 1% to more than 10%
  - Indicators of absolute mobility recently used in the literature show low bias and produce reliable rankings
  - However, some indicators (relative) with minimal bias produce high levels of re-ranking that make them uninformative to rank economies (e.g., CER050)
  - In contrast, other indicators with large bias produce more reliable rankings (e.g., IGRC)

Preview of findings

- We re-examine Emran et al. (2018)'s conclusion and offer empirical evidence against it.
  - IGPC is not always more robust than IGRC
- In addition, we show that even if the bias between both indicators is small and similar for each birth cohort separately, when we pool all of them the bias moves dramatically in favor of IGPC
- We provide novel evidence of the size of the coresidence bias for a large set of IGM indicators used in recent literature:
  - We find varying levels of coresidence bias going from less 1% to more than 10%
  - Indicators of absolute mobility recently used in the literature show low bias and produce reliable rankings
  - However, some indicators (relative) with minimal bias produce high levels of re-ranking that make them uninformative to rank economies (e.g., CER050)
  - In contrast, other indicators with large bias produce more reliable rankings (e.g., IGRC)

Preview of findings

- We re-examine Emran et al. (2018)'s conclusion and offer empirical evidence against it.
  - IGPC is not always more robust than IGRC
- In addition, we show that even if the bias between both indicators is small and similar for each birth cohort separately, when we pool all of them the bias moves dramatically in favor of IGPC
- We provide novel evidence of the size of the coresidence bias for a large set of IGM indicators used in recent literature:
  - We find varying levels of coresidence bias going from less 1% to more than 10%
  - Indicators of absolute mobility recently used in the literature show low bias and produce reliable rankings
  - However, some indicators (relative) with minimal bias produce high levels of re-ranking that make them uninformative to rank economies (e.g., CER050)
  - In contrast, other indicators with large bias produce more reliable rankings (e.g., IGRC)

Preview of findings

- We re-examine Emran et al. (2018)'s conclusion and offer empirical evidence against it.
  - IGPC is not always more robust than IGRC
- In addition, we show that even if the bias between both indicators is small and similar for each birth cohort separately, when we pool all of them the bias moves dramatically in favor of IGPC
- We provide novel evidence of the size of the coresidence bias for a large set of IGM indicators used in recent literature:
  - We find varying levels of coresidence bias going from less 1% to more than 10%
  - Indicators of absolute mobility recently used in the literature show low bias and produce reliable rankings
  - However, some indicators (relative) with minimal bias produce high levels of re-ranking that make them uninformative to rank economies (e.g., CER050)
  - In contrast, other indicators with large bias produce more reliable rankings (e.g., IGRC)

Should we prefer IGPC rather than IGRC?

Truncation bias in a Simple Model

- If a child gets married, she will leave the house; otherwise, she will stay home. The marriage decision

$$M_i = \begin{cases} 1 & if \quad v_i - wS_i^y > 0 \\ 0 & otherwise \end{cases} \tag{1}$$

- If the child is unmarried, her information is included in the survey, and the following equation holds:

$$S_i^y > \frac{v_i}{w} \equiv T_i \tag{2}$$

- Hence, the estimation of the IGRC (β) is the following linear regression equation:

$$S_i^y = \beta_0 + \beta S_i^o + \epsilon_i \quad i \in D, \quad \epsilon \sim N(0, \sigma_y^2), \quad if \quad S_i^y > T_i > 0 \tag{3}$$

Truncation bias in a Simple Model

- In the case of the IGPC (ρ), it can be written as:

$$\rho = \beta \frac{\sigma_{S^o}}{\sigma_{S^y}} \tag{4}$$

  where σ_{S^o} and σ_{S^y} are the standard deviations of years of schooling for the sample of parents and children, respectively.

- β is biased downward
  - Bias depends on correlations in the data but it is assumed downward because of an empirical regularity discussed in the literature

- $\frac{\sigma_{S^o}}{\sigma_{S^y}}$ is biased upward
  - S^y is truncated which implies lower variance
  - S^o is likely unbiased because the survey includes a random sample of parents

- Given these opposite directions, ρ is less biased than β

We re-examine the claim

- **The bias of $\frac{\sigma_{So}}{\sigma_{Sy}}$ is no necessarily upward**
  - Emran et al. (2018) assumes that β is biased downward and states that σ_{S_o} is likely to be unbiased because surveys randomly select parents (household heads and spouses)
  - However, researchers typically estimate the correlation coefficient using the set of complete cases $\rightarrow$ truncated sample of parents (in the same direction as the truncation of children given their positive association)

- **Pooling a large number of birth cohorts may favor IGPC in bias comparisons**
  - Emran et al (2018) evaluates the impact of coresidence using information of all children aged 13-60 years and subsample those who coresides with their parents
  - The approach of pooling a large number of birth cohorts favors the indicator with lower variation across cohorts (IGPC). If coresidence is random conditional on age, the benchmark weights different cohorts very differently than the coresident sample estimate just because coresidence varies with age.

The benchmark is a weighted average of the values for different cohorts

IGRC:

If there is heterogeneity across cohorts: $S^y_{ic} = \alpha_c + \beta_c S^o_{ic} + \epsilon_{ic} \qquad i \in [1, N_c] \qquad c = 1, 2$
Then the pooled estimate is a weighted average of the values of these cohorts:

$$\mathbb{E}[\hat{\beta}^{pooled}] = \beta_1 \frac{\sum_{i=1}^{N_1}(S^o_{i1} - \bar{S}^o)^2}{\sum_{i=1}^{N_1}(S^o_{i1} - \bar{S}^o)^2 + \sum_{i=1}^{N_2}(S^o_{i2} - \bar{S}^o)^2} + \beta_2 \frac{\sum_{i=1}^{N_2}(S^o_{i2} - \bar{S}^o)^2}{\sum_{i=1}^{N_1}(S^o_{i1} - \bar{S}^o)^2 + \sum_{i=1}^{N_2}(S^o_{i2} - \bar{S}^o)^2}$$

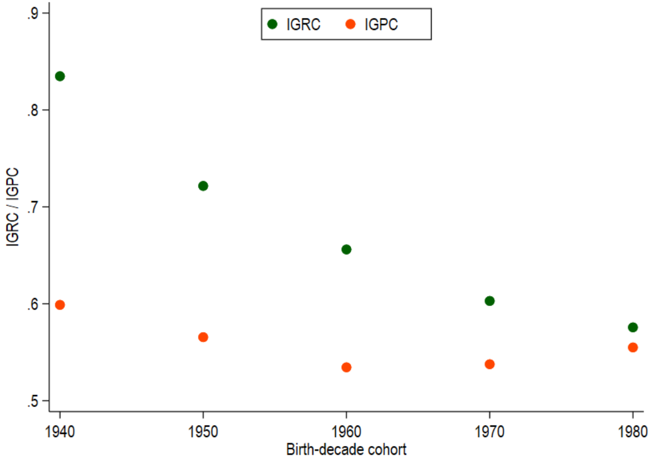
$$= \beta_1 W_1 + \beta_2 W_2$$

IGPC (same derivation using standardized years of schooling):

$$\mathbb{E}[\hat{\rho}^{pooled}] = \rho_1 \tilde{W}_1 + \rho_2 \tilde{W}_2$$

Of course, this does not matter if the parameter of interest is constant across cohorts.
However, previous evidence shows that ρ tends to be more stable than β across cohorts.

IGPC is more stable than IGRC across cohorts in India



Empirical evidence: Data

- Year 2013 wave of a national representative household survey in Colombia (Encuesta National de Calidad de Vida)

- Information about the educational attainment of all the members of each household interviewed. They are also asked about the educational attainment of their father and mother and whether they are coresiding with them

- Hence, we can compute IGM for all children and only coresidents. We do so for various birth cohorts and then we pool all of them. We compute the coresidence rate and bias for each of them

Empirical evidence: Data

- Year 2013 wave of a national representative household survey in Colombia (Encuesta National de Calidad de Vida)
- Information about the educational attainment of all the members of each household interviewed. They are also asked about the educational attainment of their father and mother and whether they are coresiding with them
- Hence, we can compute IGM for all children and only coresidents. We do so for various birth cohorts and then we pool all of them. We compute the coresidence rate and bias for each of them

Empirical evidence: Data

- Year 2013 wave of a national representative household survey in Colombia (Encuesta National de Calidad de Vida)
- Information about the educational attainment of all the members of each household interviewed. They are also asked about the educational attainment of their father and mother and whether they are coresiding with them
- Hence, we can compute IGM for all children and only coresidents. We do so for various birth cohorts and then we pool all of them. We compute the coresidence rate and bias for each of them

Empirical Evidence: Coresidence bias for two indicators of relative mobility

	Age groups (children)					
	56-65	46-55	36-45	26-35	21-25	21-65
IGRC	.69	.63	.55	.47	.39	.6
IGPC	.53	.53	.51	.53	.52	.56
IGRC (coresident sample)	.71	.61	.54	.44	.39	.51
IGPC (coresident sample)	.55	.48	.49	.5	.51	.54
Bias in IGRC (%)	3.3	-4	-1	-7.2	-.32	-15
Bias in IGPC (%)	3.7	-9	-3.9	-5.6	-.56	-3.6
Ratio of standard deviations (σ_p/σ_c)	.77	.83	.94	1.1	1.3	1.1
Ratio of SD (coresident sample)	.77	.79	.91	1.1	1.3	.94
Bias in ratio of SD (%)	.41	-5.3	-2.9	1.7	-.24	13
N	5048	7654	8598	9599	5368	35581
Coresidence rate (%)	6.5	11	17	31	53	19

Empirical Evidence: Coresidence bias for two indicators of relative mobility

	Age groups (children)					
	56-65	46-55	36-45	26-35	21-25	21-65
IGRC	.69	.63	.55	.47	.39	.6
IGPC	.53	.53	.51	.53	.52	.56
IGRC (coresident sample)	.71	.61	.54	.44	.39	.51
IGPC (coresident sample)	.55	.48	.49	.5	.51	.54
Bias in IGRC (%)	3.3	-4	-1	-7.2	-.32	-15
Bias in IGPC (%)	3.7	-9	-3.9	-5.6	-.56	-3.6
Ratio of standard deviations (σ_p/σ_c)	.77	.83	.94	1.1	1.3	1.1
Ratio of SD (coresident sample)	.77	.79	.91	1.1	1.3	.94
Bias in ratio of SD (%)	.41	-5.3	-2.9	1.7	-.24	13
N	5048	7654	8598	9599	5368	35581
Coresidence rate (%)	6.5	11	17	31	53	19

Empirical Evidence: Coresidence bias for two indicators of relative mobility

	Age groups (children)					
	56-65	46-55	36-45	26-35	21-25	21-65
IGRC	.69	.63	.55	.47	.39	.6
IGPC	.53	.53	.51	.53	.52	.56
IGRC (coresident sample)	.71	.61	.54	.44	.39	.51
IGPC (coresident sample)	.55	.48	.49	.5	.51	.54
Bias in IGRC (%)	3.3	-4	-1	-7.2	-.32	-15
Bias in IGPC (%)	3.7	-9	-3.9	-5.6	-.56	-3.6
Ratio of standard deviations (σ_p/σ_c)	.77	.83	.94	1.1	1.3	1.1
Ratio of SD (coresident sample)	.77	.79	.91	1.1	1.3	.94
Bias in ratio of SD (%)	.41	-5.3	-2.9	1.7	-.24	13
N	5048	7654	8598	9599	5368	35581
Coresidence rate (%)	6.5	11	17	31	53	19

Empirical Evidence: Coresidence bias for two indicators of relative mobility

	Age groups (children)					
	56-65	46-55	36-45	26-35	21-25	21-65
IGRC	.69	.63	.55	.47	.39	.6
IGPC	.53	.53	.51	.53	.52	.56
IGRC (coresident sample)	.71	.61	.54	.44	.39	.51
IGPC (coresident sample)	.55	.48	.49	.5	.51	.54
Bias in IGRC (%)	3.3	-4	-1	-7.2	-.32	-15
Bias in IGPC (%)	3.7	-9	-3.9	-5.6	-.56	-3.6
Ratio of standard deviations (σ_p/σ_c)	.77	.83	.94	1.1	1.3	1.1
Ratio of SD (coresident sample)	.77	.79	.91	1.1	1.3	.94
Bias in ratio of SD (%)	.41	-5.3	-2.9	1.7	-.24	13
N	5048	7654	8598	9599	5368	35581
Coresidence rate (%)	6.5	11	17	31	53	19

Coresidence bias for a large set of IGM indicators

Coresidence bias in a larger set of indicators

- We estimate IGM for the same country and (5-year) birth cohorts using two data sets:
    1. A data source containing restrospective information (individuals are asked about their parents' education) $\rightarrow$ benchmark
    2. A data source that only contains information for individuals living with their parents $\rightarrow$ coresident sample

- We compare these estimates in two dimensions:
    1. We quantify the average size of the coresidence bias (i.e., the average difference between sources as a percentage of the value computed with restropective information) for each of the 16 indicators
    2. We analyze to what extent these indicators provide valuable information to rank economies or cohorts. We compute the Spearman rank correlation between the IGM indicators using different data sources to evaluate whether the rankings are aligned

Coresidence bias in a larger set of indicators

- We estimate IGM for the same country and (5-year) birth cohorts using two data sets:
  1. A data source containing restrospective information (individuals are asked about their parents' education) $\rightarrow$ benchmark
  2. A data source that only contains information for individuals living with their parents $\rightarrow$ coresident sample

- We compare these estimates in two dimensions:
  1. We quantify the average size of the coresidence bias (i.e., the average difference between sources as a percentage of the value computed with restropective information) for each of the 16 indicators
  2. We analyze to what extent these indicators provide valuable information to rank economies or cohorts. We compute the Spearman rank correlation between the IGM indicators using different data sources to evaluate whether the rankings are aligned

Coresidence bias in a larger set of indicators

- We estimate IGM for the same country and (5-year) birth cohorts using two data sets:
    1. A data source containing restrospective information (individuals are asked about their parents' education) → benchmark
    2. A data source that only contains information for individuals living with their parents → coresident sample

- We compare these estimates in two dimensions:
    1. We quantify the average size of the coresidence bias (i.e., the average difference between sources as a percentage of the value computed with restropective information) for each of the 16 indicators
    2. We analyze to what extent these indicators provide valuable information to rank economies or cohorts. We compute the Spearman rank correlation between the IGM indicators using different data sources to evaluate whether the rankings are aligned

Coresidence bias in a larger set of indicators

- We estimate IGM for the same country and (5-year) birth cohorts using two data sets:
  1. A data source containing restrospective information (individuals are asked about their parents' education) → benchmark
  2. A data source that only contains information for individuals living with their parents → coresident sample

- We compare these estimates in two dimensions:
  1. We quantify the average size of the coresidence bias (i.e., the average difference between sources as a percentage of the value computed with restropective information) for each of the 16 indicators
  2. We analyze to what extent these indicators provide valuable information to rank economies or cohorts. We compute the Spearman rank correlation between the IGM indicators using different data sources to evaluate whether the rankings are aligned

Coresidence bias in a larger set of indicators

- We estimate IGM for the same country and (5-year) birth cohorts using two data sets:
    1. A data source containing restrospective information (individuals are asked about their parents' education) $\rightarrow$ benchmark
    2. A data source that only contains information for individuals living with their parents $\rightarrow$ coresident sample

- We compare these estimates in two dimensions:
    1. We quantify the average size of the coresidence bias (i.e., the average difference between sources as a percentage of the value computed with restropective information) for each of the 16 indicators
    2. We analyze to what extent these indicators provide valuable information to rank economies or cohorts. We compute the Spearman rank correlation between the IGM indicators using different data sources to evaluate whether the rankings are aligned

Coresidence bias in a larger set of indicators

- We estimate IGM for the same country and (5-year) birth cohorts using two data sets:
  1. A data source containing restrospective information (individuals are asked about their parents' education) $\rightarrow$ benchmark
  2. A data source that only contains information for individuals living with their parents $\rightarrow$ coresident sample

- We compare these estimates in two dimensions:
  1. We quantify the average size of the coresidence bias (i.e., the average difference between sources as a percentage of the value computed with restropective information) for each of the 16 indicators
  2. We analyze to what extent these indicators provide valuable information to rank economies or cohorts. We compute the Spearman rank correlation between the IGM indicators using different data sources to evaluate whether the rankings are aligned

Data

- 18 countries and several 5-year birth cohorts
  - Argentina, Bolivia, Brazil, Chile, Colombia, Costa Rica, Dominican Republic, Ecuador, El Salvador, Guatemala, Honduras, Mexico, Nicaragua, Panama, Peru, Uruguay, and Venezuela.
- Two data sets:
  - Latinobarometro: An opinion survey with restropective information. Individuals who were born between 1935-1995 and were at least 23 years old when they answered the survey.
  - Census data from IPUMS-International. Individuals aged 21-25 years old linked to their probable father and/or mother according to the procedures used by IPUMS for family interrelationships.

Indicators of Educational Intergenerational Mobility

Name	Description
Absolute Mobility	
YOS	Share of children with higher years of schooling than parents, $YOS = Pr(S^y > S^o \mid S^o < max(S^o))$
CAT	Share of children with higher level of education than parents, $CAT = Pr(C^y > C^o \mid C^o < max(C^o))$
MIX	A variant of CAT such that $MIX = Pr(C^y > C^o \quad or \quad C^y = C^o = max(C^o)$
BUM-primary	Bottom upward mobility: $Pr(C^y \geq primary \mid C^o < primary)$
BUM-secondary	Bottom upward mobility: $Pr(C^y \geq secondary \mid C^o < secondary)$
TDM-primary	Top down mobility: $Pr(C^y < primary \mid C^o \geq primary)$
TDM-secondary	Top down mobility: $Pr(C^y < secondary \mid C^o \geq secondary)$
UCP	Upper class persistence: $Pr(C^y \geq secondary \mid C^o \geq secondary)$
Relative mobility	
IGRC	OLS estimate of the slope (β) in $S^y = \alpha + \beta S^o$
IGPC	Pearson correlation coefficient (ρ), where $\rho = Corr(S^o, S^y)$
IGSC	Spearman correlation coefficient, $IGSC = Corr(R^y, R^o)$
CER050	Expected rank of children with parents in bottom half, $CER050 = \mathbb{E}(R^y \mid R^o \leq 50)$
BHQ4	Prob. of reaching top quartile if parents are in bottom half, $BHQ4 = Pr(R^y > 75 \mid R^o \leq 50)$
Movement	
M1	Average change in schooling between generations, $M1 = \frac{1}{N} \sum \lvert S_i^y - S_i^o \rvert$
M2	Average directional change in schooling between generations, $M2 = \frac{1}{N} \sum (S_i^y - S_i^o)$
DIF	Same as M2 but for children with parents that did not complete tertiary

Main results: All children vs. coresidents

Indicator	Average difference (%)	Rank correlation
Absolute mobility		
UCP	0.693	0.551
BUM-primary	-2.199	0.910
YOS	-2.959	0.718
TDM-secondary	12.844	0.551
TDM-primary	14.705	0.737
BUM-secondary	-17.127	0.855
CAT	-30.847	0.744
MIX	-30.951	0.702
Relative mobility		
CER050	6.361	0.186
IGPC	10.854	0.490
IGSC	12.448	0.368
IGRC	18.817	0.820
BHQ4	40.174	0.164
Movement		
M1	-10.812	0.766
M2	-12.159	0.747
DIF	-13.032	0.799

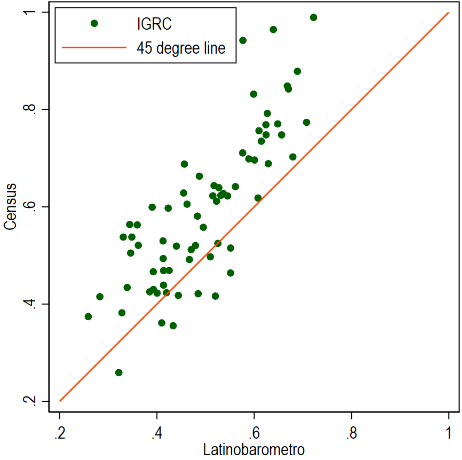
Main results: All children vs. coresidents

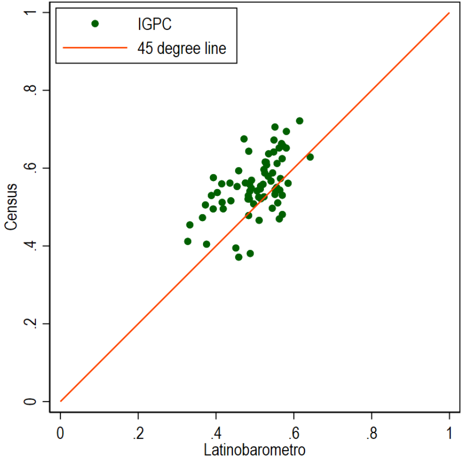
Indicator	Average difference (%)	Rank correlation
Absolute mobility		
UCP	0.693	0.551
BUM-primary	-2.199	0.910
YOS	-2.959	0.718
TDM-secondary	12.844	0.551
TDM-primary	14.705	0.737
BUM-secondary	-17.127	0.855
CAT	-30.847	0.744
MIX	-30.951	0.702
Relative mobility		
CER050	6.361	0.186
IGPC	10.854	0.490
IGSC	12.448	0.368
IGRC	18.817	0.820
BHQ4	40.174	0.164
Movement		
M1	-10.812	0.766
M2	-12.159	0.747
DIF	-13.032	0.799

Main results: All children vs. coresidents

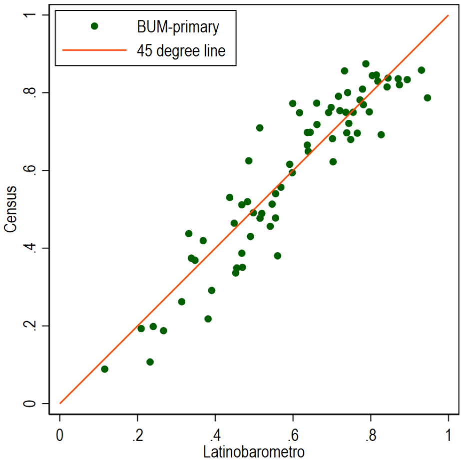
Indicator	Average difference (%)	Rank correlation
Absolute mobility		
UCP	0.693	0.551
BUM-primary	-2.199	0.910
YOS	-2.959	0.718
TDM-secondary	12.844	0.551
TDM-primary	14.705	0.737
BUM-secondary	-17.127	0.855
CAT	-30.847	0.744
MIX	-30.951	0.702
Relative mobility		
CER050	6.361	0.186
IGPC	10.854	0.490
IGSC	12.448	0.368
IGRC	18.817	0.820
BHQ4	40.174	0.164
Movement		
M1	-10.812	0.766
M2	-12.159	0.747
DIF	-13.032	0.799

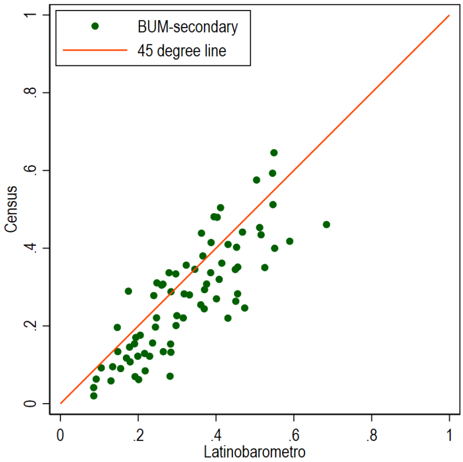
Relative mobility: All children vs. coresidents



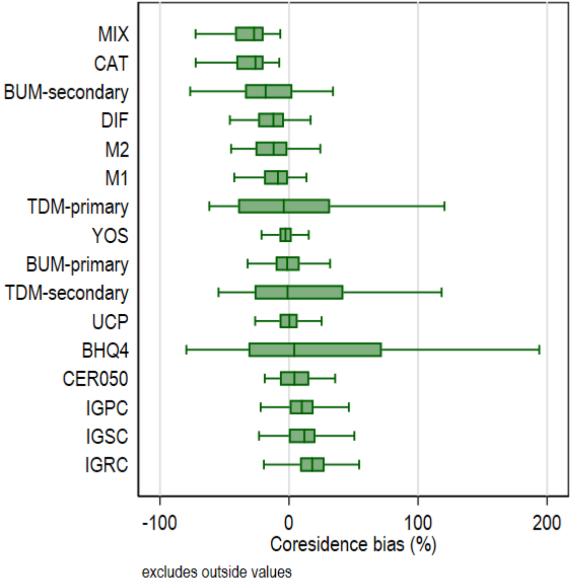


Absolute mobility: All children vs. coresidents

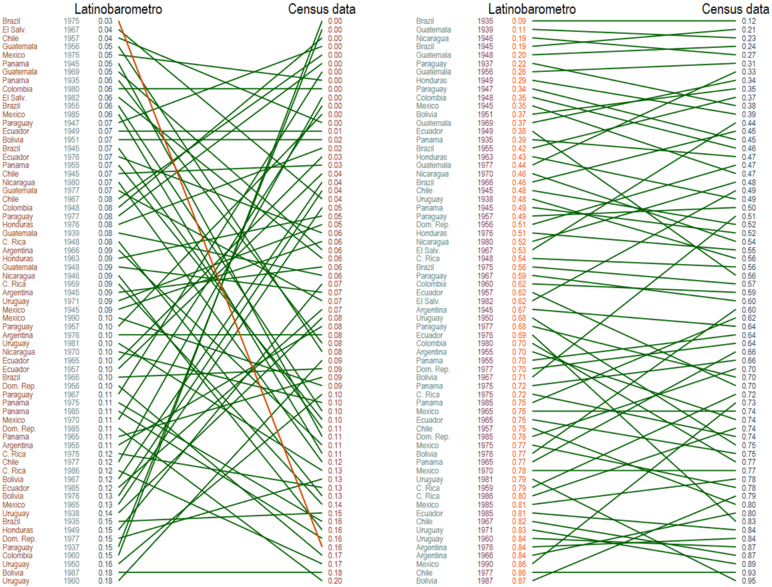




Variability of the coresidence bias



Re-ranking: BHQ4 (corr. = 0.16) and BUM-primary (corr. = 0.91)



Re-ranking and discrepancies happen even in absence of coresidence

Indicator	Average difference (%)	Rank correlation
Absolute mobility		
BUM-secondary	-1.985	0.840
UCP	3.639	0.518
Relative mobility		
IGSC	3.642	0.067
IGPC	7.019	0.050
IGRC	13.210	0.699
Movement		
M2	-0.438	0.590
M1	-0.961	0.638

Note: This table compares 113 estimates with household surveys to those with Latinobarometro social survey.

Re-ranking and discrepancies happen even in absence of coresidence

Indicator	Average difference (%)	Rank correlation
Absolute mobility		
BUM-secondary	-1.985	0.840
UCP	3.639	0.518
Relative mobility		
IGSC	3.642	0.067
IGPC	7.019	0.050
IGRC	13.210	0.699
Movement		
M2	-0.438	0.590
M1	-0.961	0.638

Note: This table compares 113 estimates with household surveys to those with Latinobarometro social survey.

Summary

- In this paper we contribute to the understanding of intergenerational mobility in education by studing the impact of coresidence bias on its measurement

- IGRC vs. IGPC:
  - We show that IGPC and IGRC are not always biased downward. IGPC is on average less (upward) biased than IGRC. However, despite its lower bias, it provides a less reliable ranking

- Large set of indicators:
  - Absolute IGM
    - Recently used indicators are relatively robust: low bias and high rank correlation
  - Relative IGM
    - Larger bias than the lowest biased absolute indicators. Some indicators with low bias show low rank correlation while other with higher bias show higher rank correlation

Thank you