How biased are measures of intergenerational educational mobility? Practical guidance from a simulation framework

Marcelo Cardona & Sam Jones

18 August 2021

Overview

- Focus is on the measurement of intergenerational (educational) mobility
- Bewildering range of empirical choices face applied researchers
- Evidence on bias associated with different choices is scattered and incomplete
- Address this gap using a flexible simulation framework
- Provide some practical guidance

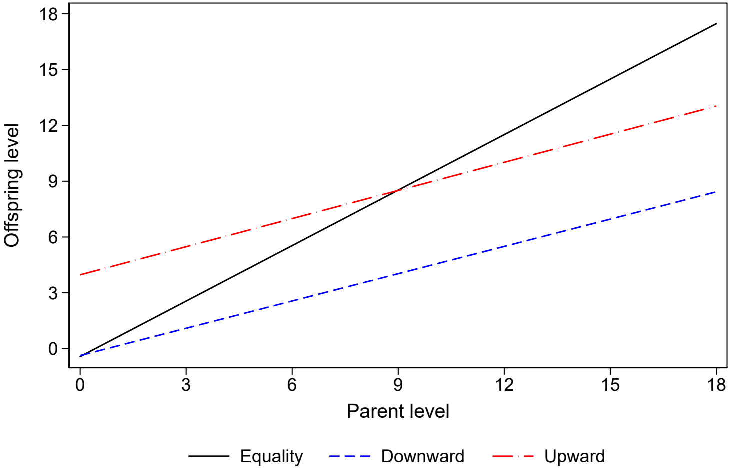
Agenda

(1) Measuring mobility

Broad terrain

- Broad sense, IGM refers to the degree to which advantages and disadvantages of individuals persist across generations (e.g., Great Gatsby curve) – dimension of social justice
- Relevant in multiple domains, but education often a primary focus in developing countries (as robust predictor of well-being and measured directly)
- No single accepted definition: “IGM is a complex concept and may mean different things to different people ... A natural consequence ... is that there is no consensus on how IGM should be measured, so many indices are available to an applied researcher.” (Savegnago, 2016: 386)

Stylized linear example



Multiple empirical metrics

Define: c = child outcome; p = parental outcome

Class	Focus	Example
Heritability	Slope	Correl(c,p)
Performance	Distance	E($c \mid p$) – p
	Transition	$\Pr(c > p \mid p \leq p^*)$
Distribution	Inequality	Gini($c + p$) / Gini(p)

Distinctions between:

- Parametric (regression) vs. non-parametric approaches (e.g., transition matrices)
- Number of free parameters (to be estimated)

Multiple data transforms

Monotonic data transformations often used to remove (some) nuisance parameters and facilitate comparison/interpretation

Transform	Mean	Variance
None	μ	σ
Variance stabilization	μ	1
Mean-variance stabilization	0	1
Rank	1/2	1/12

Issues:

- Applied separately or jointly?
- Sensitivity to reference distribution(s) & outliers

Examples of data transforms

Choice of both mobility metric (class) **and** data transform defines what particular aspect of mobility will be captured.

... and sensitivity to different forms of measurement error.

Examples:

		Transformation	
Measure	Exp.	None (levels)	Rank-Rank
Heritability	$\beta^{\cdot}$	$\text{Cov}(c^*, p^*)/\text{Var}(p^*)$	$12 \cdot \text{Cov}(c^{*r}, p^{*r})$
Out-perform.	$\Delta^{\cdot}_{p_{25}}$	$\text{E}(c^*) - \beta\left(\text{E}(p^*) - p^*_{25}\right) - p^*_{25}$	$.25(1 - \beta^r)$
Upward mob.	$\text{up}^{\cdot}$	$\text{E}(c^*_i > p^*_i \mid p^*_i \leq p_{50})$	$\text{E}(c^{*q}_i > p^{*q}_i \mid p^{*q}_i \leq .5)$

Empirical challenges

Few parent-child matched administrative datasets exist (outside Scandinavia).

Co-resident, self-reported data is commonplace (e.g., census).

Three generic challenges $\rightarrow$ corrections:

Problem	Poss. corrections	Example
Incomplete education	Sample restrictions	Child $\geq$ 18
Reporting error	Ranking	Percentile
Missing co-residents	Prediction, bounding	$c - \epsilon \leq p \leq c + \epsilon$

Bias characterization

Key questions:

1. how material are these biases (in plausible contexts)?
2. do they vary across measures (metrics $\times$ transforms)
3. do conventional empirical corrections help?

Difficult to answer analytically:

- Different sources of bias *may* offset one another [see paper]
- Various measures are non-linear (in expectation)

So, let's run some scenarios!

(2) Simulation framework

Basic set-up

Extension of generalized errors-in-variables approach (c.f, Nybom and Stuhler, 2017), to account for systematic missing observations and incomplete education:

$$c_{ij}^* = \alpha_j \bar{p}^* + \beta_j p_i^* + \theta_0 \varepsilon_{0ij} \tag{1a}$$

$$c_{ij} = c_0 + (1 + \lambda_c) c_{ij}^* - \delta m_i^* + \nu_i \tag{1b}$$

$$p_i = p_0 + (1 + \lambda_p) p_i^* + \mu_i \tag{1c}$$

$$\nu_i = \theta_1 \varepsilon_{1i}, \ \mu_i = \gamma \nu_i + \theta_2 \varepsilon_{2i} \tag{1d}$$

$$\varepsilon_{1ij} \sim \mathcal{N}_l^u(0, \sqrt{1 - \beta_j^2} \cdot \sigma_{p^*}), \ \varepsilon_{1i}, \varepsilon_{2i}, \sim \mathcal{N}_l^u(0, 1) \tag{1e}$$

where *i* are individuals and *j* indexes seen vs unseen groups; $\mathcal{N}_l^u$ is a truncated normal distribution; and $\theta_0, \theta_1, \theta_2 \geq 0$ are scaling parameters.

Metrics × transforms × corrections

Metric	Expression	Description
Non-Herit.	$1-\beta$	Reverse of slope
Out-perform.	$\hat{\alpha}+p_{25}(\hat{\beta}-1)$	Exp. diff at 25th pc
Inequality	$1-\mathrm{g}(c+p)/\mathrm{g}(c)$	Fall in inequality
Move sh.	$N^{-1}\sum_N \mathbf{1}(\lvert c-p\rvert>\pi)$	Share moved by $>\pi$
Move sh. (+)	$N^{-1}\sum_N \mathbf{1}(c-p>\pi)$	Share moved upward
Move sh. (++)	$N^{-1}\sum_N \mathbf{1}(c-p>\pi)\mid p<p_{50}$	Conditional share moved up
Wgt. move sh.	$N^{-1}\sum_N \lvert c-p\rvert/\max(c-p)$	Weighted share moved
Wgt. move sh. (+)	as above $\cdot\mathbf{1}(c^r>p^r)$	Weighted share moved up
Wgt. move sh. (++)	as above $\mid p<p_{50}$	Conditional share moved up

Metrics × transforms × corrections

Transform	Description
None	-
Var.	Variance stabilization (separate)
Ref.	Percentiles of parental distribution
Rank	Percentiles (separate)

Correction	Description
None	-
Enroll	Only children NOT enrolled in school
Age	Only children with age such that $c > p$
Predict	Predict missing p from c (age-corrected)

Implementation steps

1. Calibrate parameters of true DGP (e.g., p^*, α_j, β_j)
2. Set assumptions for measurement error structure (e.g., $\theta_1, \theta_2, p_0, c_0, \lambda_c, \lambda_p, \gamma, \delta$)
3. Draw stochastic variables (error terms, iid)
4. Simulate full dataset
5. Impose corrections (to observed data) and transforms
6. Calculate mobility metrics (true vs observed)

Note: step 6 can be bootstrapped to verify error distributions.

Measures of bias

Here for β, but can be any chosen final metric:

$$\text{MB}(\beta) = \frac{1}{N}\sum_{n=1}^{N}(\hat{\beta}_n - \beta_n) \tag{2a}$$

$$\text{MFB}(\beta) = \frac{1}{N}\sum_{n=1}^{N}\frac{2(\hat{\beta}_n - \beta_n)}{(\mid \hat{\beta}_n \mid + \mid \beta_n \mid)} \tag{2b}$$

$$\text{MAFB}(\beta) = \frac{1}{N}\sum_{n=1}^{N}\frac{2(\mid \hat{\beta}_n - \beta_n \mid)}{(\mid \hat{\beta}_n \mid + \mid \beta_n \mid)} \tag{2c}$$

where $n = \{1, \ldots, N\}$ indexes simulation iterations (for a given scenario).

(3) Results: stylized low education case

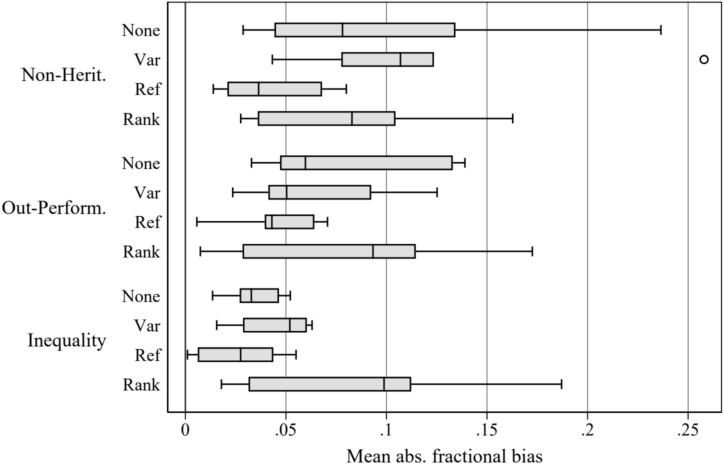
Stylized low education case

Calibrate model to Mozambique and consider 6 distinct measurement error scenarios.

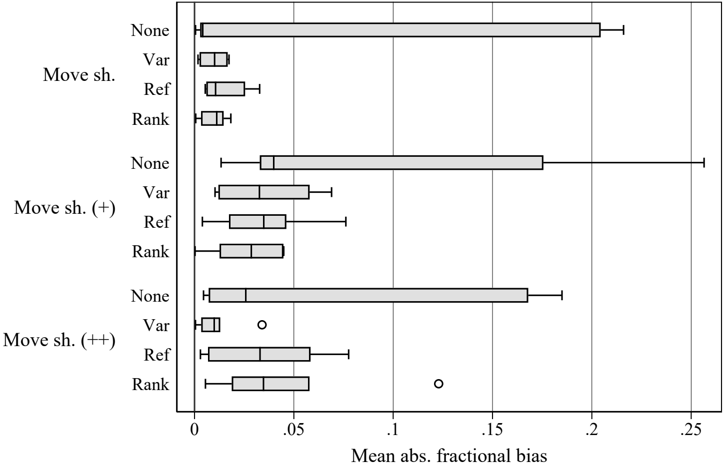
Table: Mean outcomes (all 6 scenarios $\times$50 iterations = 300 obs.)

	True	Estimated bias (no correction)			
Metric $\downarrow$ // Transform $\rightarrow$	None	None	Var	Ref.	Rank
Non-Herit.	0.60	0.05	0.06	0.03	0.04
Out-Perform.	0.17	0.01	0.01	-0.01	0.01
Inequality	0.54	-0.00	-0.00	-0.01	0.02
Move sh.	0.89	-0.06	0.00	0.01	0.00
Move sh. (+)	0.61	-0.05	-0.02	-0.02	-0.00
Move sh. (++)	0.79	-0.05	-0.01	-0.03	-0.01
Wgt. m. sh.	0.22	0.00	0.01	0.00	0.01
Wgt. m. sh. (+)	0.16	-0.01	0.00	-0.01	0.00
Wgt. m. sh. (++)	0.22	-0.00	0.01	-0.02	-0.00

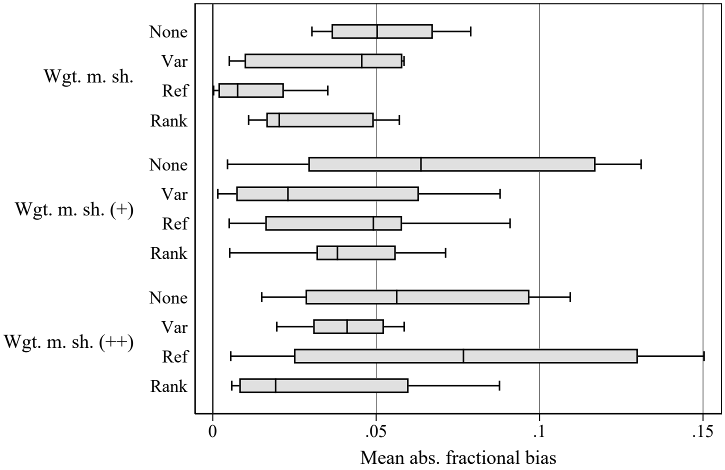
All scenarios, no corrections (metrics 1-3)



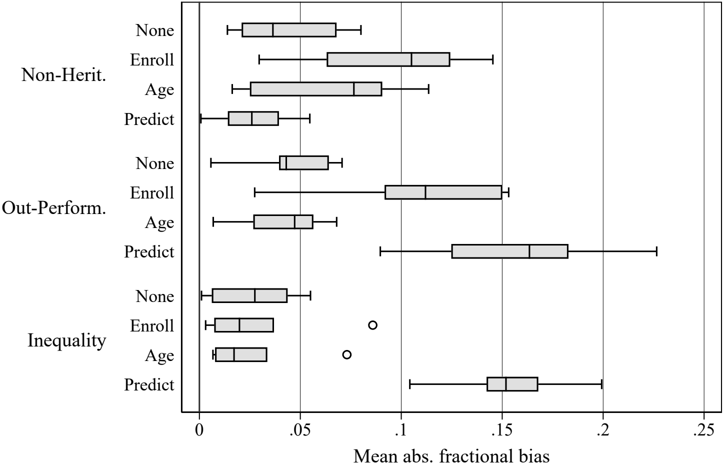
All scenarios, no corrections (metrics 4-6)



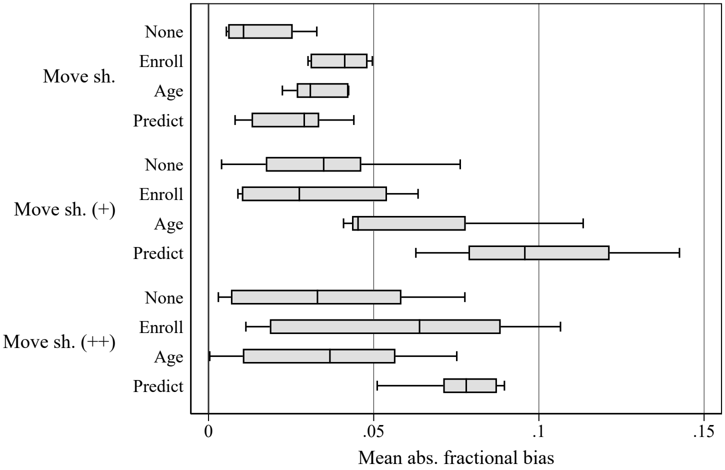
All scenarios, no corrections (metrics 7-9)



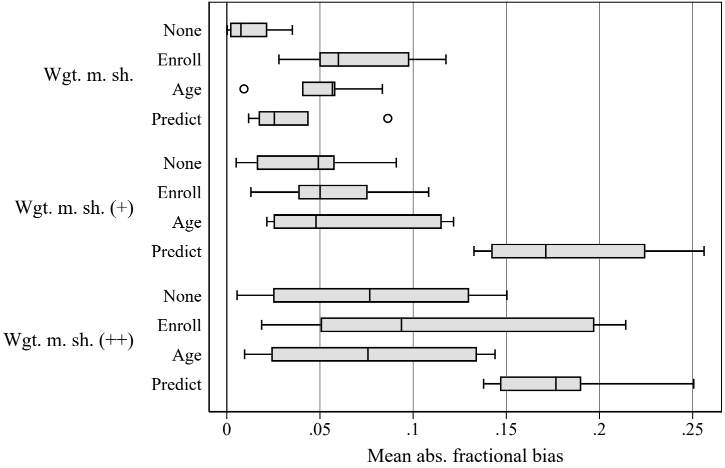
All scenarios, ref. transform (metrics 1-3)



All scenarios, ref. transform (metrics 4-6)



All scenarios, ref. transform (metrics 7-9)



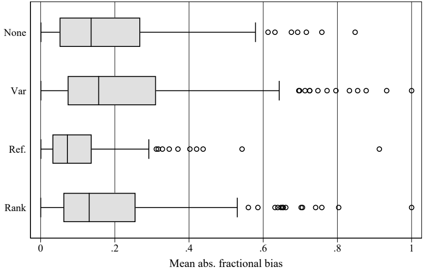
(4) Results: generalized case

Generalized scenarios, no corrections

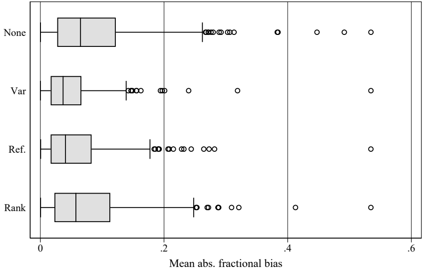
Run 300 separate draws of *both* DGP and measurement error structure (jointly).

On each draw look at bias associated with all combinations of transforms and corrections.

Non-heritability



Move sh. (++)



	Non-herit		Out-Perf.		Move % (++)	
Constant	0.150***	0.150***	-0.065***	-0.065***	-0.030***	-0.030***
	(0.004)	(0.004)	(0.006)	(0.006)	(0.003)	(0.003)
Var. trans.	0.033***	0.033***	0.088***	0.088***	0.022***	0.022***
	(0.005)	(0.005)	(0.006)	(0.006)	(0.003)	(0.003)
$\times\ \alpha$		-0.035***		-0.037***		0.006*
		(0.006)		(0.008)		(0.003)
$\times\ \beta$		0.003		-0.013**		-0.002
		(0.006)		(0.006)		(0.003)
$\times\ \bar{p}$		-0.012**		0.057***		0.022***
		(0.005)		(0.008)		(0.003)
Ref. trans.	-0.051***	-0.051***	0.011	0.011*	0.002	0.002
	(0.004)	(0.003)	(0.006)	(0.006)	(0.003)	(0.003)
$\times\ \alpha$		-0.045***		0.023**		0.014***
		(0.004)		(0.009)		(0.003)
$\times\ \beta$		-0.046***		-0.001		0.001
		(0.004)		(0.006)		(0.003)
$\times\ \bar{p}$		0.027***		0.015*		-0.002
		(0.004)		(0.009)		(0.003)
Rank. trans.	0.055***	0.055***	0.232***	0.232***	0.081***	0.081***
	(0.004)	(0.004)	(0.008)	(0.008)	(0.004)	(0.004)
$\times\ \alpha$		-0.001		0.032***		0.019***
		(0.005)		(0.011)		(0.004)
$\times\ \beta$		0.025***		0.046***		0.018***
		(0.005)		(0.008)		(0.004)
$\times\ \bar{p}$		0.004		0.007		-0.015***
		(0.004)		(0.010)		(0.004)
Enroll corr.	0.082***	0.082***	-0.114***	-0.115***	-0.043***	-0.043***
	(0.006)	(0.006)	(0.010)	(0.010)	(0.005)	(0.005)
Age corr.	-0.037***	-0.037***	0.023***	0.023***	0.004	0.004
	(0.005)	(0.005)	(0.007)	(0.007)	(0.003)	(0.003)
Fit corr.	-0.020***	-0.020***	-0.076***	-0.076***	-0.027***	-0.027***
	(0.006)	(0.006)	(0.009)	(0.009)	(0.004)	(0.004)
Obs.	4,800	4,800	4,800	4,800	4,800	4,800
R2 (adj.)	0.644	0.681	0.598	0.615	0.516	0.537
RMSE	0.107	0.102	0.172	0.168	0.075	0.073

	Non-herit		Out-Perf.		Move % (++)	
Constant	0.160***	0.160***	0.227***	0.227***	0.093***	0.093***
	(0.003)	(0.003)	(0.005)	(0.005)	(0.002)	(0.002)
Var. trans.	0.047***	0.047***	-0.104***	-0.104***	-0.043***	-0.043***
	(0.004)	(0.004)	(0.005)	(0.005)	(0.002)	(0.002)
$\times\ \alpha$		-0.010**		0.045***		0.010***
		(0.005)		(0.006)		(0.003)
$\times\ \beta$		0.016***		0.023***		0.003
		(0.005)		(0.005)		(0.002)
$\times\ \bar{p}$		-0.006		-0.037***		-0.022***
		(0.004)		(0.006)		(0.003)
Ref. trans.	-0.061***	-0.061***	-0.048***	-0.048***	-0.031***	-0.031***
	(0.003)	(0.003)	(0.005)	(0.005)	(0.002)	(0.002)
$\times\ \alpha$		-0.043***		0.003		0.005**
		(0.004)		(0.007)		(0.003)
$\times\ \beta$		-0.051***		-0.002		-0.006***
		(0.004)		(0.005)		(0.002)
$\times\ \bar{p}$		0.025***		0.007		-0.003
		(0.004)		(0.007)		(0.003)
Rank. trans.	0.042***	0.042***	-0.032***	-0.032***	-0.015***	-0.015***
	(0.004)	(0.003)	(0.007)	(0.006)	(0.003)	(0.003)
$\times\ \alpha$		-0.000		0.117***		0.035***
		(0.004)		(0.008)		(0.003)
$\times\ \beta$		0.018***		0.107***		0.018***
		(0.004)		(0.006)		(0.003)
$\times\ \bar{p}$		0.003		-0.056***		-0.027***
		(0.004)		(0.007)		(0.003)
Enroll corr.	0.075***	0.075***	0.028***	0.028***	0.019***	0.019***
	(0.005)	(0.005)	(0.009)	(0.008)	(0.003)	(0.003)
Age corr.	-0.014***	-0.014***	-0.009	-0.009*	-0.003	-0.003
	(0.004)	(0.004)	(0.006)	(0.005)	(0.002)	(0.002)
Fit corr.	-0.014***	-0.014***	0.048***	0.048***	0.015***	0.015***
	(0.005)	(0.004)	(0.008)	(0.007)	(0.003)	(0.003)
Obs.	4,800	4,800	4,800	4,800	4,800	4,800
R2 (adj.)	0.713	0.752	0.393	0.504	0.428	0.475
RMSE	0.089	0.083	0.145	0.131	0.058	0.055

(5) Summary

Summary

No single empirical approach to measuring IG mobility.

Multiple choices: **metrics** $\times$ **transforms** $\times$ **corrections**

Lack of evidence on sensitivity to measurement error & systematic missing observations.

We developed a flexible simulation framework to quantify bias for educational attainment.

Considered stylized scenarios (for a given country) and generalized scenarios (wide range of parameter combinations)

Main lessons

1. Non-heritability coefficient $(1 - \beta)$ upward biased
2. Lower bias of non-parametric measures (e.g., % up)
3. Data transforms *often* reduce bias (mean & absolute)
4. Additional corrections do not generally reduce bias
5. BUT no single "always-lowest-bias" measure – characteristics of the case matter (shapes of true CDFs)

Recommendation

Use a *calibrated* simulation model to estimate upper/lower bounds on bias for a plausible range of 'true' models.

e.g., 95% of time bias is within $|\,\varepsilon\,|$.

Practical example – working on it!