Supplementary Material

Inequality of Opportunity in South Asia: The Puzzle of Educational Gains Without Consumption Gains

Maurizio Bussolo* Vito Peragine† Fabian Reutzel‡

Published in *Oxford Open Economics*

*World Bank, mbussolo@worldbank.org
†Department of Economics and Finance, University of Bari, vitorocco.peragine@uniba.it
‡Paris School of Economics, Université Paris 1 Panthéon-Sorbonne, fabian.reutzel@psemail.eu

A.1 Data

In the following, we explain the survey selection across countries and outcomes summarized in table A1. Whenever possible, we rely on World Bank harmonized versions of the surveys discussed below: the SARMD database for household (HH) surveys and the GLD database for labor force (LF) surveys. When both sources are available for a given country, we prefer the national labor force survey (LFS) over the household survey for labor market outcomes.

Afghanistan. We draw on multiple waves of nationally representative surveys to maximize the time span and, hence, the range of birth cohorts covered: the Integrated Expenditure and Labor Force Survey (IELFS; 2019), its predecessor the Afghanistan Living Conditions Survey (ALCS; 2013-14, 2016-17), and the earlier National Risk and Vulnerability Assessment (NRVA; 2007-08, 2011-12). Pooling these surveys is admissible for the educational analysis given their representativeness, but the consumption measures are not comparable across waves. To validate the pooled cohort-based analysis of labor market outcomes, we examine cross-survey alignment and find that: (i) LFP appears slightly overstated relative to ILO estimates (figure A16), with the 2007 wave particularly affected (68% vs. 48%) and therefore excluded; (ii) wage-employment shares remain stable across waves; and (iii) wages can only be measured in 2008, 2014, and 2020, and while the associated Gini aligns across waves, the limited sample (<1,000 observations per cohort; see table A2) prohibits a cohort-based analysis.

Bangladesh. We employ all available waves of the Household Income and Expenditure Survey (HIES; 2000, 2005, 2010, 2016, 2022) for the education and consumption outcomes, the comprehensive nationally representative survey conducted by the Bangladesh Bureau of Statistics (BBS) for poverty measurement.[32] For the labor market analysis, we prefer the LFS: while its LFP is slightly elevated relative to ILO estimates (68% vs. 60% in 2005), the HIES departs more severely (47% in 2005). The LFS also appears to oversample economically active women (e.g., in 2016: 41% in LFS, 14% in HIES, and 33% in the ILO estimate), whereas the HIES substantially overreports wage employment, particularly for women (e.g., 74% in HIES vs. 41% in LFS for women, and 60% vs. 41% for men); see figures A22 and A23. The slight reduction observed in LFS-based LFP IOp therefore represents an upper bound on the true improvement, while the HIES-based wage-employment estimates can be disregarded given their severe measurement issues. We also considered drawing on the Bangladesh Integrated Household Survey (BIHS) but decided against it, as its representativeness is limited to the rural population.

Bhutan. We employ the Bhutan Living Standards Survey (BLSS; 2003, 2007, 2012, 2017, 2022) for all outcome dimensions, conducted by the National Statistics Bureau (NSB), which collects information on education, poverty, employment, and happiness, among other topics. For the labor market, no wage income is recorded, and we exclude the 2007 wave because its spike in LFP strongly distorts the cohort-based estimates. Given the absence of labor market information in the 2022 BLSS, we revert to the Labor Force Survey conducted in the same year, though we exclude this wave from the cohort-based analysis given its structurally higher LFP rate. Due to the short time span of available LFS

[32]The HIES 2022 is excluded from the pooled cohort-based consumption analysis due to a different estimation methodology, and from the LFP robustness checks given a surge in reported LFP.

waves (2018-2022, provided by the SARLAB database) and, hence, the limited pseudo-panel dimension, we prefer the BLSS cohort-based outcomes. For consumption, the 2003 and 2022 waves differ from the others in the estimation procedure of the consumption aggregate, which precludes their inclusion in the pooled cohort-based analysis.

India. We rely on the Household Consumption Expenditure Surveys conducted by the National Sample Survey Office (NSSO; 1993, 2004, 2009, 2011, 2022) as provided by SARMD. In the appendix, we additionally leverage the India Human Development Survey (IHDS; 2005, 2011) for a basic assessment of the importance of parental background, as it directly records this information (section A.4). For consumption, the recall period changed in 2022 from a uniform recall (past 7 days) to a "modified mixed reference period (MMRP)" approach. Although both recall periods were elicited in 2011, allowing a comparison across methodologies, additional changes in consumption items and sample design, together with the shift from a single-visit to a three-module, three-visit design, impede comparability between 2022 and earlier waves and motivate the exclusion of the former. For the labor market, two potential sources are available: the household surveys (NSS 1993-2011, HCES 2022) and the LFS, comprising the Periodic Labour Force Survey (PLFS, 2017-2022) and its predecessor the Employment and Unemployment Survey (EUS, 1987-2011). Comparing the two, we find that: (i) NSS and LFS align for LFP, both in terms of population shares and IOp across cohorts (table A3); (ii) a change in the sampling and reporting of LFP in the PLFS (2017-22) leads us to restrict those waves to the younger cohorts (1975-79, 1980-84) not covered in earlier survey waves, thereby preventing the mixing of imperfectly comparable sources (Chaudhary and Thakur, 2025); (iii) the LFS's harmonized paid-employment definition is broader than the wage-employment definition recovered from the NSS (45% vs. 28% in 2011 for 35-54 year olds); and (iv) the wage analysis can only be conducted on the LFS given the absence of wage reporting in the consumption-focused NSS waves we draw on.

Nepal. Education estimates are based on the 2011 National Population and Housing Census (NPHC) and the 2022 Nepal Living Standards Survey (NLSS).[33] For consumption, we draw on the NLSS (1995, 2003, 2011, 2022), which follows the World Bank's Living Standards Measurement Survey methodology, but its consumption estimates are comparable only for 1995 and 2003, precluding more recent waves from the pooled cohort-based analysis. For the labor market, three potential LFS waves are available (1998, 2008, 2017), but we note that: (i) the reported LFP of 87% in the first two waves is more than double the ILO estimate (42%) and therefore not truly representative under the international definition; and (ii) only the 2017 wave resolves this issue (LFP of 40%), but relying on a single wave precludes our pseudo-panel technique. We therefore turn to the NLSS (1995, 2003, 2011) also for the labor market analysis. Given its overstated LFP (approximately 70%), we present results only for wage employment as a share of the full population, which we also check for alignment with the LFS 1998 and 2008 (figure A17).[34] A cohort-based wage IOp is not feasible given the limited sample size (table A2).

[33]The NLSS relies on sampling weights for national representativeness, whereas the NPHC does not require these given its census sample size. Hence, for all cohorts covered by the NPHC, we disregard the NLSS and rely on the latter only for the young cohorts (1990-94, 1995-00) not covered by the census. Estimates based on earlier waves of the NLSS align with the NPHC-based estimates.

[34]Both the LFP and wage-employment definitions in the 2022 NLSS differ from previous waves, precluding its inclusion in the pooled cohort-based analysis.

Pakistan. We draw on the Pakistan Social and Living Standards Measurement Survey (PSLM; 2010, 2012, 2014, 2019) and the Household Integrated Economic Survey (HIES; 2007, 2010, 2011, 2013, 2015, 2018) conducted by the Pakistan Bureau of Statistics (PBS). Earlier waves cannot be included due to the absence of our demographic circumstance variable (i.e., language of the interview). The PSLM is collected with representativeness at the district level, whereas the HIES sample is representative only at the national/provincial level; the PSLM sample is therefore significantly larger than that of the HIES, and the two surveys were conducted in alternating years. Since we only rely on the higher geographic level (province) of residence as a circumstance, both surveys are pertinent to our analysis. We additionally leverage the Pakistan Integrated Household Survey (PIHS, 1991) to cover older cohorts (born in the 1950s) in the education analysis and provide in the appendix a basic assessment of the importance of parental background, given its directly recorded question on this matter (section A.4). For the labor market dimension, we again rely on the HIES, as the LFS does not record a demographic circumstance variable. We exclude the 2013 wave from the analysis of LFP and wage-employment, as its LFP diverges from previous and subsequent waves (figure A16).

Sri Lanka. We rely on the Household Income and Expenditure Survey (HIES; 1990, 1995, 2002, 2006, 2009, 2012, 2016, 2019) for the educational and consumption dimensions, the country's main data source for poverty analysis, conducted by the Department of Census and Statistics (DCS). The 1990 and 1995 consumption estimates differ in their estimation methodology and are therefore excluded from the pooled cohort-based analysis. For the labor market dimension, we draw on the country's bi-/annual LFS (1992-2021), which aligns closely with the HIES in terms of LFP and wage levels; the share of wage employment appears 8-10 percentage points lower in the LFS, yet IOp estimates fully align across the two sources (table A3). A discrepancy emerges for wage IOp in the oldest cohorts (1950s), which can be reconciled by the different survey coverage periods: the LFS observes these individuals already in the 1990s and early 2000s, whereas harmonized wage data in the HIES is only available from 2006 onward.

Table A1: Survey Overview by Outcome

Country	Survey	Education	Consumption	LFP & Wage-Employment	Wages
Afghanistan	NRVA	2008, 2012		2012	
Afghanistan	ALCS	2014, 2017	(**2017**)	2014, 2017	
Afghanistan	IELFS	**2020**		**2020**	(**2020**)
Bangladesh	HIES	2000, 2005, 2010, 2016, **2022**	2000, 2005, 2010, 2016, (**2022**)	2000, 2005, 2010, 2016	2005, 2010, 2016
Bangladesh	LFS			2000*, 2005, 2010, 2016, (**2022**)	2005, 2010, 2013, 2015, 2016, (**2022**)
Bhutan	BLSS	2003, 2007, 2012, 2017, **2022**	2003, 2007, 2012, 2017, (**2022**)	2003, 2012, 2017	
Bhutan	LFS			2018, 2019, 2020, (**2022**)	
India	IHDS	2005, 2011			
India	NSS	1993, 2004, 2009, 2011	1993, 2004, 2009, 2011	1993, 2004, 2009, 2011	
India	HCES	(**2022**)	(**2022**)		
India	EUS			1983, 1987, 1993, 1999, 2004, 2005, 2007, 2009, 2011	1987, 1993, 2004, 2005, 2007, 2009, 2011
India	PLFS			2017, 2018, 2019, 2020, 2021, **2022**	2017, 2018, 2019, 2020, 2021, **2022**
Nepal	NLSS	**2022**	1995, 2003, (2010), (**2022**)	1995, 2003, 2010, (2022)	
Nepal	NPHC	2011			
Nepal	LFS			1998, 2008, (**2017**)	(**2017**)
Pakistan	HIES	2007, 2010, 2011, 2013, 2015, 2018	2007, 2010, 2011, 2013, 2015, **2018**	2007, 2010, 2011, 2015, **2018**	2010, 2011, 2015, **2018**
Pakistan	PIHS	1991			
Pakistan	PSLM	2010, 2012, 2014, **2019**			
Sri Lanka	HIES	1990, 1995, 2002, 2006, 2009, 2012, 2016, **2019**	2002, 2006, 2009, 2012, 2016 **2019**	1990, 1995, 2002, 2006, 2009, 2012, 2016, 2019	2006, 2009, 2012, 2016, 2019
Sri Lanka	LFS			1992, 1993, 1994, 1995, 1996, 1998, 1999, 2000, 2001, 2002, 2003, 2006, 2007, 2008, 2011, 2012, 2013, 2014, 2015, 2019, 2020, **2021**	1992, 1993, 1994, 1995, 1998, 1999, 2000, 2001, 2002, 2003, 2006, 2007, 2008, 2011, 2012, 2013, 2014, 2015, 2019, 2020, **2021**

Notes: See above for detailed description of survey selection. Survey waves in bold indicate the data source for the cross-sectional inequality and IOp estimates presented in table A3. Survey waves in brackets indicate their non-comparability to other waves and, hence, their exclusion from the cohort analysis. For Nepal, the survey wave used for LFP and wage-employment in the cross-sectional analysis is the LFS 2017 but this survey cannot be used for the cohort analysis as definitions of LFP and wage-employment change compared to previous LFS waves (1998, 2008). As this earlier LFP definition is not internationally comparable (see figure A16), those waves can only be used for wage-employment in the cohort analysis and sample size of both, the LFS and the NLSS, is too small to perform the latter for wages. For Bangladesh, the employment definition of the 2000 LFS (indicated by *) does not align with subsequent years so it is excluded from the wage-employment analysis.
Source: Own calculations based on harmonized survey data (table A1).

Table A2: Sample Size by Birth Cohort

		Education	Consumption	LFP		Wage Employment		Wages	
Country	Cohort	HH Survey	HH Survey	HH Survey	LF Survey	HH Survey	LF Survey	HH Survey	LF Survey
Afghanistan	1950-54	11933							
	1955-59	14307							
	1960-64	20101		10646		5261			
	1965-69	22535		15259		8029			
	1970-74	28813		21197		11387			
	1975-79	32503		22554		12235			
	1980-84	41662		10994		5256			
	1985-89	41957							
	1990-94	31372							
	1995-00	12381							
Bangladesh	1950-54	10957	2696						
	1955-59	14005	5578	5573	15137	2012	9413	1047	3253
	1960-64	17877	11587	11602	32172	4891	21105	2610	6902
	1965-69	19994	16126	16157	45915	7698	30790	4310	10617
	1970-74	24461	15705	15748	46231	8321	30950	4870	9621
	1975-79	24593	12371	12432	34601	6478	23316	4157	7586
	1980-84	27907			18304		12087		2501
	1985-89	23720							
	1990-94	20627							
	1995-00	4978							
Bhutan	1950-54	7216	1424						
	1955-59	9539	3702						
	1960-64	10273	5763	3531		2700			
	1965-69	12465	8085	5395	5447	4170	4675		
	1970-74	14093	6620	4706	7688	3601	6732		
	1975-79	16304	4844	4829	8761	3686	7713		
	1980-84	19033			11251		9792		
	1985-89	17203							
	1990-94	13231							
	1995-00	7434							
India	1950-54	201313	54871	54330	126982	36557	86554		24147
	1955-59	257337	95962	94701	181492	63427	122785		41348
	1960-64	309046	86241	85406	179544	57019	121194		39566
	1965-69	363083	102636	102559	169358	67957	114705		49172
	1970-74	339985	67256	67207	102638	42856	68362		31027
	1975-79	320575			200318		133802		59138
	1980-84	318635			181777		119364		55461
	1985-89	226580							
	1990-94	98732							
	1995-00	97677							
Nepal	1950-54	127042				1076			
	1955-59	135066				1965			
	1960-64	168667				2015			
	1965-69	194186				1695			
	1970-74	225799							
	1975-79	271946							
	1980-84	290236							
	1985-89	349803							
	1990-94	2821							
	1995-00	2985							
Pakistan	1950-54	67294							
	1955-59	86723	8722	8185		4129		1085	
	1960-64	109705	20359	16196		8983		3732	
	1965-69	132553	32693	28043		15796		7114	
	1970-74	150012	33130	28015		16346		8327	
	1975-79	165392	26195	20828		11995		7213	
	1980-84	198583	10936	10936		6436		3867	
	1985-89	215998							
	1990-94	130825							
	1995-00	49156							
Sri Lanka	1950-54	36731	7424	19170	45221	9436	29986	994	14833
	1955-59	40240	15751	24058	55271	12885	37299	4382	18905
	1960-64	45621	24530	26268	64693	13526	44734	7521	23181
	1965-69	44165	28559	28557	58879	16772	40430	10494	21036
	1970-74	37493	22735	22733	48724	15126	33294	9561	16653
	1975-79	33081	13853	13853	36374	9273	24425	5843	12187
	1980-84	30766	7715	7715	18034	5014	11896	3318	5260
	1985-89	18270							
	1990-94	10618							
	1995-00	2702							

Notes: The table depicts the sample size available for each cohort by outcome and data source. The first row of the header indicates the outcome and its second row the data source, i.e., household (HH) surveys from the SARMD database and labor force (LF) surveys from the GLD database, both pre-harmonized by the World Bank. See above for the explanation of the survey selection.
Source: Own calculations based on harmonized survey data (table A1).

Table A3: Total Inequality & Relative IOp across Outcomes & Cohorts

		Education		Consumption		LFP				Wage Employment				Wages			
						HH Survey		LF Survey		HH Survey		LF Survey		HH Survey		LF Survey	
Country	Cohort	Gini	IOp	Gini	IOp	Share	IOp	Share	IOp	Share	IOp	Share	IOp	Gini	IOp	Gini	IOp
Afghanistan	1950-54	0.88	0.53														
	1955-59	0.86	0.55														
	1960-64	0.85	0.58			0.58	0.30			0.15	0.43						
	1965-69	0.84	0.58			0.61	0.26			0.16	0.36						
	1970-74	0.85	0.58			0.62	0.27			0.16	0.32						
	1975-79	0.84	0.58			0.63	0.26			0.16	0.29						
	1980-84	0.82	0.58			0.62	0.25			0.18	0.28						
	1985-89	0.77	0.56														
	1990-94	0.69	0.54														
	1995-00	0.62	0.50														
Bangladesh	1950-54	0.74	0.46	0.35	0.42												
	1955-59	0.71	0.45	0.36	0.40	0.57	0.37	0.63	0.29	0.52	0.09	0.36	0.12	0.44	0.53	0.42	0.46
	1960-64	0.70	0.43	0.37	0.38	0.55	0.39	0.66	0.24	0.55	0.07	0.34	0.12	0.44	0.49	0.38	0.36
	1965-69	0.68	0.40	0.36	0.37	0.54	0.39	0.68	0.21	0.57	0.06	0.36	0.12	0.41	0.49	0.37	0.33
	1970-74	0.64	0.37	0.33	0.36	0.55	0.38	0.68	0.21	0.60	0.05	0.37	0.13	0.38	0.49	0.34	0.28
	1975-79	0.58	0.31	0.32	0.37	0.55	0.38	0.69	0.20	0.66	0.04	0.40	0.11	0.37	0.48	0.29	0.32
	1980-84	0.49	0.26					0.67	0.22			0.46	0.10			0.29	0.25
	1985-89	0.40	0.20														
	1990-94	0.35	0.18														
	1995-00	0.28	0.17														
Bhutan	1950-54	0.92	0.62	0.42	0.50												
	1955-59	0.91	0.62	0.40	0.48												
	1960-64	0.88	0.64	0.40	0.46	0.77	0.08			0.33	0.34						
	1965-69	0.83	0.61	0.39	0.47	0.78	0.08	0.86	0.05	0.35	0.34	0.24	0.39				
	1970-74	0.77	0.57	0.38	0.46	0.78	0.10	0.88	0.05	0.42	0.28	0.28	0.37				
	1975-79	0.72	0.52	0.37	0.43	0.77	0.10	0.89	0.05	0.42	0.28	0.34	0.30				
	1980-84	0.62	0.43					0.89	0.05			0.40	0.25				
	1985-89	0.51	0.40														
	1990-94	0.39	0.33														
	1995-00	0.24	0.30														
India	1950-54	0.65	0.58	0.37	0.47	0.68	0.21	0.68	0.20	0.26	0.37	0.45	0.13			0.57	0.61
	1955-59	0.62	0.56	0.37	0.47	0.67	0.22	0.68	0.20	0.25	0.36	0.44	0.13			0.58	0.63
	1960-64	0.60	0.55	0.39	0.51	0.67	0.23	0.68	0.21	0.26	0.35	0.45	0.12			0.57	0.64
	1965-69	0.57	0.52	0.38	0.51	0.67	0.23	0.68	0.20	0.24	0.35	0.44	0.13			0.54	0.60
	1970-74	0.51	0.51	0.38	0.50	0.64	0.25	0.67	0.22	0.25	0.33	0.47	0.12			0.52	0.58
	1975-79	0.46	0.49					0.67	0.22			0.46	0.12			0.46	0.54
	1980-84	0.42	0.48					0.66	0.22			0.48	0.11			0.44	0.54
	1985-89	0.37	0.45														
	1990-94	0.32	0.45														
	1995-00	0.25	0.42														
Nepal	1950-54	0.83	0.57							0.22	0.22						
	1955-59	0.77	0.57							0.23	0.23						
	1960-64	0.75	0.58							0.24	0.20						
	1965-69	0.69	0.58							0.27	0.25						
	1970-74	0.64	0.57														
	1975-79	0.58	0.54														
	1980-84	0.50	0.52														
	1985-89	0.42	0.49														
	1990-94	0.32	0.47														
	1995-00	0.25	0.44														
Pakistan	1950-54	0.74	0.57														
	1955-59	0.71	0.56	0.34	0.37	0.52	0.33			0.54	0.13			0.49	0.56		
	1960-64	0.69	0.55	0.34	0.38	0.56	0.31			0.57	0.12			0.45	0.54		
	1965-69	0.66	0.53	0.34	0.40	0.57	0.30			0.57	0.11			0.44	0.58		
	1970-74	0.62	0.51	0.33	0.41	0.59	0.29			0.60	0.10			0.42	0.56		
	1975-79	0.57	0.48	0.33	0.42	0.58	0.29			0.60	0.09			0.41	0.61		
	1980-84	0.53	0.46	0.32	0.46	0.60	0.30			0.60	0.08			0.39	0.65		
	1985-89	0.50	0.44														
	1990-94	0.49	0.42														
	1995-00	0.46	0.41														
Sri Lanka	1950-54	0.28	0.34	0.41	0.32	0.68	0.19	0.68	0.18	0.64	0.07	0.56	0.10	0.41	0.40	0.39	0.36
	1955-59	0.27	0.32	0.41	0.30	0.68	0.18	0.69	0.17	0.63	0.08	0.56	0.09	0.42	0.37	0.40	0.34
	1960-64	0.26	0.31	0.40	0.31	0.68	0.18	0.70	0.16	0.62	0.06	0.55	0.08	0.38	0.35	0.38	0.38
	1965-69	0.23	0.30	0.40	0.31	0.68	0.19	0.70	0.17	0.63	0.06	0.56	0.07	0.38	0.38	0.39	0.39
	1970-74	0.20	0.32	0.39	0.33	0.69	0.19	0.69	0.17	0.64	0.05	0.56	0.06	0.38	0.38	0.38	0.41
	1975-79	0.16	0.32	0.39	0.31	0.68	0.19	0.68	0.18	0.63	0.05	0.58	0.06	0.36	0.41	0.36	0.47
	1980-84	0.14	0.32	0.39	0.33	0.66	0.21	0.68	0.19	0.67	0.04	0.61	0.05	0.33	0.42	0.36	0.60
	1985-89	0.13	0.33														
	1990-94	0.11	0.31														
	1995-00	0.09	0.35														

Notes: Given the sample size displayed in table A2, wages are excluded from the cohort based IOp analysis for Afghanistan. The latest cohort of Nepal is excluded as it only observed in one cross-section (no pseudo panel).
Source: Own calculations based on harmonized survey data (table A1).

Table A4: Circumstances - Geographical Region

Country	Levels		Names	
	Region	Sub-Region	Region	Sub-Region
Afghanistan	7	34	UN Subregions	Provinces (wilåyat)
Bangladesh	8	64	Divisions	Districts
Bhutan	4	20	Zones (dzongdeys)	Districts (dzongkhags)
India	6	27	geographic	States
Nepal	5	7	Regions	Provinces
Pakistan	4	-	Provinces	-
Sri Lanka	7	9	geographic	Provinces

Notes: Throughout the analysis the region level is used for the circumstance definition as the number of regions is more similar across countries (i.e., more "comparable" IOp estimates given a similar number of regressors). We check for sensitivity of the IOp estimates with respect to geographical granularity. For India, we follow Singh (2012) in grouping the 26 states and 1 union territory (Delhi; administrative organization prior to 2000) into 6 geographical regions. For Pakistan, no consistent higher regional granularity can be recovered due to multiple changes in administrative structure.

Table A5: Geo-spatial Migration

			Migration Birth to Current Location			
Country	Survey	Year	Urban	Region	Sub-Region	N
Afghanistan	ALCS	2014	6.68	3.18	7.33	80759
Afghanistan	ALCS	2017		3.64	6.51	76669
India	IHDS	2005	12.50		4.86	97001
India	IHDS	2011	9.44		3.78	98786
Nepal	NLSS	1995	8.61			10714
Nepal	NLSS	2010	22.02	9.65	27.48	17573
Nepal	NPHC	2011	2.09	6.61	18.59	2598164
Pakistan	PSLM	2019	3.29	1.37	8.12	528840

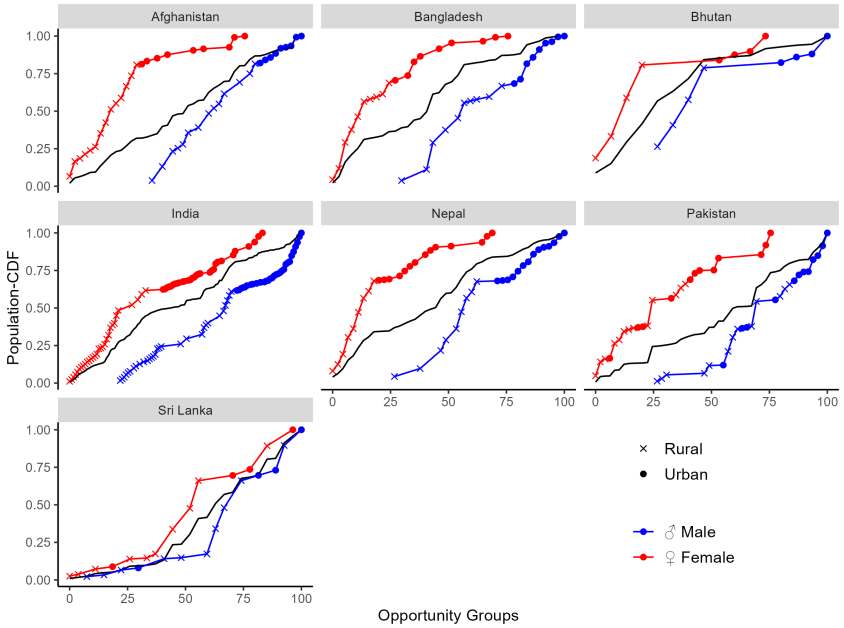
Notes: The table displays the share of individuals in % that migrated from birth location to the current location by different geographical granularity. The sample is restricted to individuals aged at least 18 as these are the youngest individuals considered in the IOp estimations. For Nepal, the NLSS used for the IOp in consumption estimates, systematically oversamples urban population compared to NPHC (35.09% vs. 31.28%) which translates into substantially higher share of individuals that report urban migration. For India, the IHDS does not report the individual's location of birth but asks in a binary nature whether the individual was born in a different state. Hence, migration across states (i.e., sub-region) is an upper bound for mobility on the regional level. Further, mobility estimates are in line with evidence for low male migration flows due to caste-based rural insurance networks (Munshi and Rosenzweig, 2016).
Source: Own calculations based on harmonized survey data (table A1).

Table A6: Circumstances - Demographic Group

Country	Definition	Levels	Labels
Afghanistan	Ethnicity	5	Pashtun; Tajik; Uzbek; Nuristani; Mixed Area + Others
Bangladesh	Religion	2	Muslim; Hindu + Others
Bhutan	-		
India	Caste/Religion	5	Scheduled Caste; Scheduled Tribe; Other Backward Class; Muslim; Forward + Others
Nepal	Caste/Religion	4	Janajati; Khas; Muslim; Dalit + Others
Pakistan	Language	5	Urdu; Punjabi + Hindko; Sindhi; Pushtu; Others
Sri Lanka	Ethnicity	3	Sri Lanka Tamil/Moors; Indian Tamil; Sinhalese + Others

Notes: For Afghanistan, the variable is constructed based on the largest demographic group in the province of residence reported in the Asia Foundation's "Survey for Afghan people 2006–2019". If a group has a population share of more than 70 percent in a given region, the individual residing there is attributed to this demographic group. If there is no such clear majority group the individual is assigned to the residual other category. In this manner, five demographic groups are formed: Pashtun, Tajik, Uzbek, Nuristani, Others + Mixed Area. In Bangladesh, a binary classification in Muslim (majority population) and non-Muslim is utilized. For Bhutan, no demographic/ethnic group variable could be found across surveys. For India, caste and religion are combined into a salient classification of five groups: Scheduled Caste, Scheduled Tribe, Other Backward Class, Muslim and Others. For Nepal, a similar compound classification of caste and religion is derived: Janajati, Khas, Muslim and Dalit + Others. For Pakistan, the language of the interview is used for classifying demographic groups: Urdu, Punjabi/Hindko, Sindhi, Pushtu and Others. For Sri Lanka, ethnicity is directly reported and renders distinction into four groups: Sinhalese + Others, Sri Lanka Tamil, Indian Tamil and Sri Lanka Moors.

Figure A1: Population CDF - Years of Education



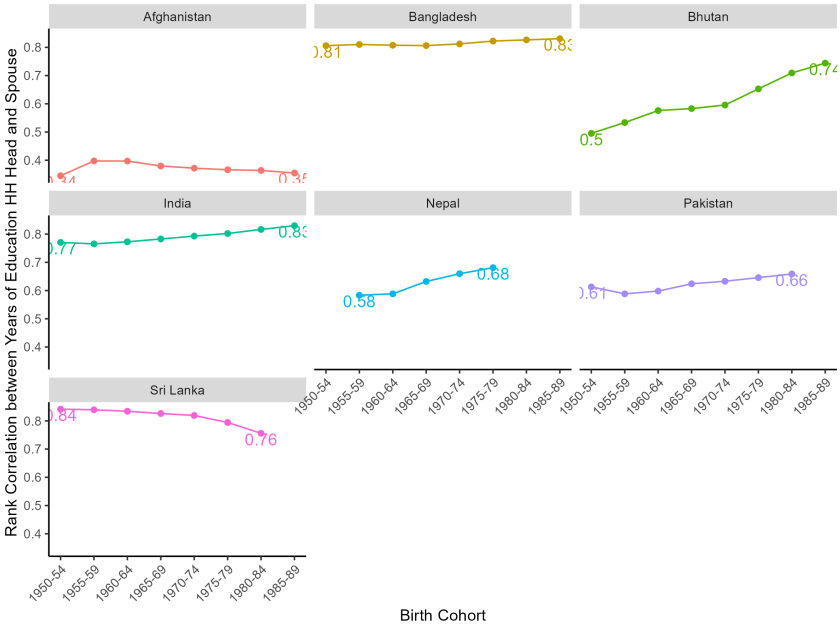
Notes: The figure depicts the Cumulative Distribution Function (CDF) of the female (red) and male (blue) population across countries and opportunity groups in the 1950s cohort, i.e., the proportion of the gender-specific population with an predicted educational attainment smaller than or equal to opportunity group indicated on the x-axis. Group characteristics (circumstances) are indicated by the symbol (Urbanity) and its color (Gender), e.g., red crosses indicate females living in rural areas (see figure A4 for further explanation). We note, that females are overrepresented in the bottom half of the population with respect to educational attainment, i.e., the female population CDF surpasses 0.5 well before its male analogue and the bottom have to be almost exclusively composed of individuals living in rural areas.
Source: Own calculations based on harmonized survey data (table A1).

Table A7: Population Summary Opportunity Profiles

Country	Cohort	P25	Median	P75	P90	P95	P90P50	P95P50
Afghanistan	1950-54	0.66	1.58	2.80	3.85	4.17	2.44	2.64
	1995-00	3.49	5.87	8.02	9.59	10.52	1.63	1.79
Bangladesh	1950-54	2.59	3.55	4.96	6.23	6.47	1.76	1.82
	1995-00	8.73	9.38	10.14	10.62	10.99	1.13	1.17
Bhutan	1950-54	0.63	1.61	2.59	3.29	3.42	2.04	2.12
	1995-00	8.66	9.68	11.04	11.25	11.29	1.16	1.17
India	1950-54	2.56	3.83	6.25	6.96	8.15	1.82	2.13
	1995-00	9.44	10.87	12.21	13.60	14.18	1.25	1.30
Nepal	1950-54	0.89	1.91	3.52	4.51	4.87	2.36	2.55
	1995-00	6.64	8.32	10.03	11.11	11.59	1.33	1.39
Pakistan	1950-54	0.45	2.49	4.18	5.12	5.65	2.06	2.27
	1995-00	4.25	5.83	7.39	8.59	9.21	1.47	1.58
Sri Lanka	1950-54	4.98	6.35	7.68	8.68	9.05	1.37	1.42
	1995-00	10.24	10.87	11.38	11.80	12.05	1.08	1.11

Notes: The table summarizes the predicted educational attainment as display in figure A4 weighted by population shares along the displayed percentiles of the population and reports their ratios.
Source: Own calculations based on harmonized survey data (table A1).

Figure A2: Assortative Mating across Birth Cohorts



Notes: The figures depicts the Spearman rank correlation coefficient for years of education between spouses across cohorts and countries.
Source: Own calculations based on harmonized survey data (table A1).

A.2 Convexity of Returns to Education

This annex provides the detailed econometric analysis supporting the convexity findings summarized in section 5.2. We estimate returns to education across three sequential labor market outcomes—labor force participation, wage employment, and wages—to identify where in the labor market structure educational credentials matter and whether these returns exhibit convexity.

Estimation Strategy. We estimate returns to education by regressing labor market outcomes on years of education (with quadratic term), age controls, and circumstances, allowing all coefficients to vary by birth cohort. Formally, for country c and birth cohort b:

$$\begin{aligned}\text{outcome}_{icb} = \sum_b D_b \Big(\alpha_{cb} + \beta^1_{cb}\text{educ}_{icb} + \beta^2_{cb}\text{educ}^2_{icb} \\ + \delta_{cb}\text{Circ}_{icb} \Big) + \gamma^1_{cb}\text{age}_{icb} + \gamma^2_{cb}\text{age}^2_{icb} + \epsilon_{icb}\end{aligned} \tag{4}$$

where i indexes individuals, D_b are cohort dummy variables, and all regressors, except age polynomials, are fully interacted with cohort dummies such that each coefficient varies by cohort. The oldest cohort serves as the baseline, and regressions are pooled across cohorts within each country.[35] For labor force participation and wage employment (binary outcomes), we estimate equation (3) via a linear probability model regression. For wages, we use OLS on log wages. The quadratic specification allows for non-linear returns, capturing potential convexity.

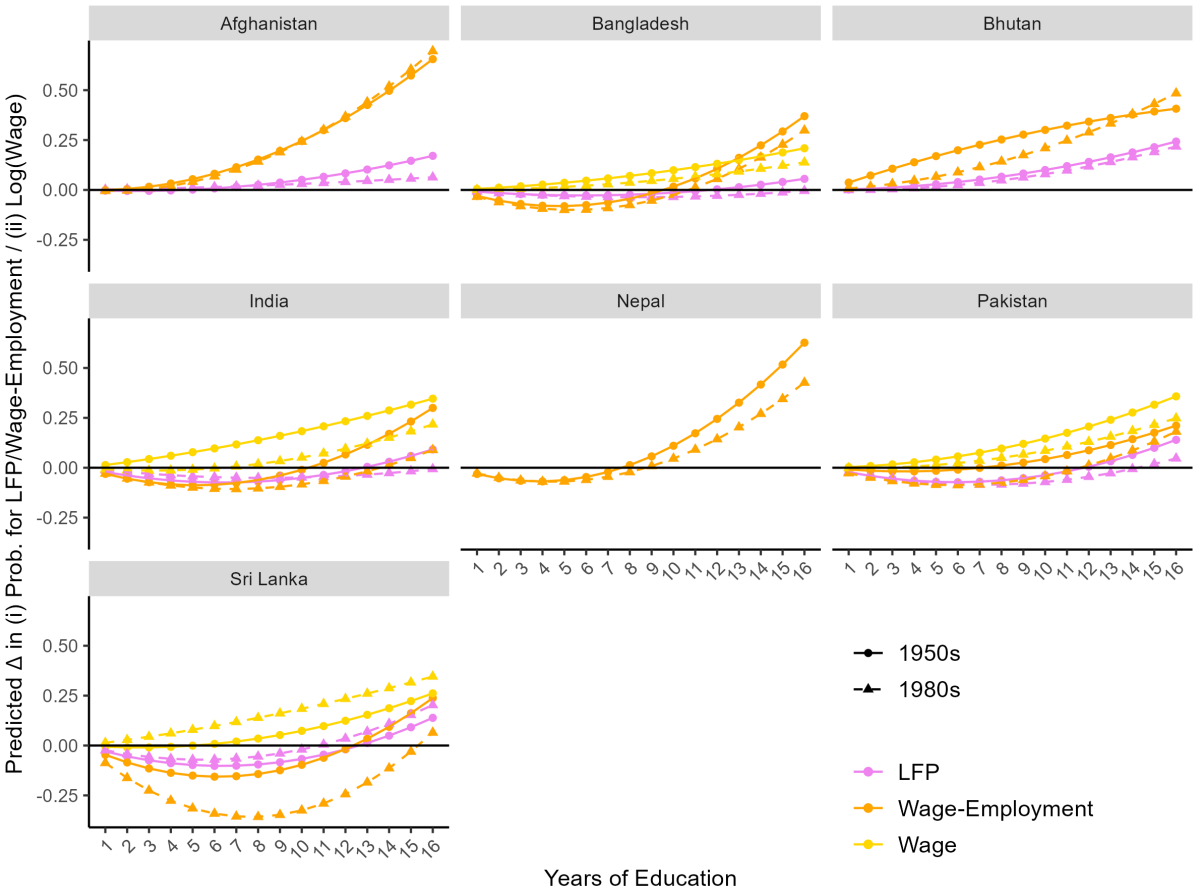
Returns to education across labor market stages. Figure A3 presents the predicted returns to education by years of educational attainment for the oldest (1950s) and youngest (1980s) cohorts across outcomes. The figure plots the predicted change in outcomes for individuals with different years of education compared to individuals with no formal education, holding other variables constant.[36]

We examine convexity across all three labor market outcomes—participation, wage employment, and wages—because educational barriers may operate sequentially through the labor market structure. If educational expansion concentrated at levels with low labor market returns while access to high-return education remained restricted for disadvantaged groups, this pattern should manifest not only in wage determination but also in selection into labor force participation and access to formal employment. A finding of convex returns only at the wage stage would suggest education equalizes market entry but not earnings conditional on entry. Conversely, convexity across all three outcomes would indicate that marginalized groups gained access primarily to education levels insufficient for overcoming barriers at any stage—from initial participation through formal employment to earnings.

[35] For testing the significance of coefficient changes across cohorts, we estimate the pooled specification with all regressors interacted with cohort dummies and test the interaction coefficients. Complete regression results are reported in tables A8 to A10.

[36] For labor force participation (LFP, violet) and wage employment (orange), predictions represent differential probabilities of participation/employment. For wages (yellow), the predicted change in log(wages) is rescaled by the cohort's median wage level for visual comparability across countries. Complete regression results appear in table A8 (LFP), table A9 (wage employment), and table A10 (wages)

Figure A3: Labor Market Returns to Education



Notes: The figure depicts returns to education by years of educational attainment, i.e., predicted changes in outcomes for different years of education compared to individuals without formal education, for the oldest (1950s) and youngest (1980s) cohorts across outcomes (see tables A8 to A10 for regression results). For labor force participation (LFP) and wage employment this corresponds to the differential probability of participation/employment. For wages, the predicted change in log(wages) has rescaled by the cohorts median wage level for reasons of visual comparability.
Source: Own calculations based on harmonized survey data (table A1).

Several patterns emerge. First, returns exhibit substantial convexity across all three outcomes in most countries, though the extent varies considerably. For labor force participation, educational returns are limited and often materialize only at higher educational levels. In several countries (Afghanistan, Bangladesh, Pakistan), individuals with 5-8 years of education show little to no higher probability of labor force participation compared to those with no education. Only at 10+ years does a positive association emerge, and even then it remains modest in most countries except Bhutan and Sri Lanka in recent cohorts.

For wage employment, convexity appears more pronounced, particularly in Afghanistan and Nepal, which are characterized by especially low wage employment shares (below 25% of the labor force; see table A3). In these settings, wage employment appears nearly unattainable for those with less than primary education—in Afghanistan and Bhutan for the 1950s cohort, the predicted probability of wage employment is more than 50 percentage points higher for those with upper secondary education than for those with less than primary education, while the latter group has only a 7-9% baseline probability of wage employment.

For wages, convexity in the 1950s cohorts appears more limited in Bangladesh and India, while it is more pronounced in Pakistan and Sri Lanka. This cross-country variation may reflect differences in labor market structure and the prevalence of formal sector employment (see table A3 for cross-country comparisons).

Second, returns to education generally diminished across cohorts while simultaneously exhibiting slight intensification of convexity. This pattern—declining levels but steeper curvature—appears across most countries and outcomes. The exception is Sri Lanka, which exhibits slightly increasing and less convex returns to education for the youngest cohort, a pattern that coincides with that country's higher and increasing wage IOp (figure 4). Further, a notable association between education and labor force participation emerges only in the latest cohorts for Bhutan and Sri Lanka, countries that differ substantially in their development trajectories and labor market structures.

The South Asian pattern thus differs from the case studies examined by Bourguignon et al. (2005) in important respects: (i) educational expansion focused strongly on providing basic education to all, precisely the range of education with limited labor market returns, and (ii) declining levels of returns have curtailed potential inequality-augmenting effects that might otherwise have resulted from expanded educational opportunities. Sri Lanka represents the case closest to Bourguignon's framework—overall educational expansion (not an expansion concentrated at lower levels, given the coverage of basic education already substantial in early cohorts, figure 5) combined with convex (and slightly increasing) returns to education for wages (figure A3, yellow lines, with the 1980s line above that of the 1950s)—yet it experienced declining rather than rising wage inequality across cohorts (table A3). However, Bourguignon's framework focuses on returns to earnings conditional on employment, without considering access to employment itself. When we examine returns to education for accessing wage employment (figure A3, orange lines), we observe substantially lower returns in the 1980s cohort: wage jobs became less education-dependent. This weaker selection mechanism may have contributed to reducing overall wage inequality through a composition effect—expanding the pool of wage workers to include more medium-educated individuals—even as returns to education for wages slightly increased.[37]

These findings suggest that the combination of bottom-concentrated educational expansion (section 5.1) and convex returns substantially limits the potential for educational gains to translate into labor market gains. However, even these limited returns might have reduced consumption IOp had they accrued equally across types. Section 5.3 examines whether labor markets reward equally educated individuals equally, or whether circumstances continue to predict labor market outcomes even controlling for educational attainment.

[37] Clearly, Sri Lanka's declining wage inequality may be driven by additional factors such as wage compression policies, sectoral shifts, and other labor market institutions that may have operated alongside these educational dynamics.

Table A8: Labor Market Returns to Education: LFP

	Afghanistan				Bangladesh				Bhutan				India				Pakistan				Sri Lanka			
	1960-64		1980-84		1955-59		1980-84		1960-64		1975-79		1950-54		1980-84		1955-59		1980-84		1950-54		1980-84	
Variable	Coef.	p-value	Coef.	p-value	Coef.	p-value	Coef.	p-value	Coef.	p-value	Coef.	p-value	Coef.	p-value	Coef.	p-value	Coef.	p-value	Coef.	p-value	Coef.	p-value	Coef.	p-value
(Intercept (differential Cohort FE))	0.303**	0.019	−0.093***	0.000	−0.316***	0.000	0.047**	0.027	0.381	0.120	−0.022	0.272	0.427***	0.000	0.000	0.940	0.289***	0.003	0.074***	0.000	0.372***	0.000	−0.025	0.354
Education	−0.004	0.296	0.006	0.299	−0.009***	0.000	0.001	0.864	0.001	0.823	−0.003	0.692	−0.023***	0.000	0.011***	0.000	−0.025***	0.000	0.001	0.870	−0.032***	0.000	0.006	0.241
Education^2	0.001***	0.004	−0.001*	0.069	0.001***	0.000	0.000	0.261	0.001*	0.070	0.000	0.877	0.002***	0.000	−0.001***	0.000	0.002***	0.000	0.000	0.206	0.003***	0.000	0.000	0.640
Age	0.037***	0.000			0.064***	0.000			0.032***	0.005			0.031***	0.000			0.032***	0.000			0.030***	0.000		
Age^2	−0.001***	0.000			−0.001***	0.000			0.000***	0.001			0.000***	0.000			0.000***	0.000			0.000***	0.000		
Female	−0.649***	0.000	0.038***	0.006	−0.727***	0.000	0.197***	0.000	−0.262***	0.000	−0.050***	0.006	−0.571***	0.000	−0.018***	0.000	−0.689***	0.000	−0.028**	0.026	−0.497***	0.000	−0.035***	0.000
Urban	−0.036***	0.004	−0.030*	0.067	−0.021***	0.003	−0.027**	0.022	−0.205***	0.000	0.043*	0.057	−0.106***	0.000	0.033***	0.000	−0.071***	0.000	0.014	0.289	−0.050***	0.000	0.051***	0.000
Dem. Group2	0.016**	0.047			−0.035***	0.000							0.060***	0.000			0.012***	0.008			−0.072***	0.000		
Dem. Group3	0.152***	0.000											−0.010***	0.000			0.080***	0.000			0.092***	0.000		
Dem. Group4	0.127***	0.000											−0.089***	0.000			−0.083***	0.000						
Dem. Group5	0.005	0.472											−0.042***	0.000			0.124***	0.000						
Region2	−0.027	0.114	0.151***	0.000	−0.015	0.241	0.010	0.674	0.039*	0.059	0.024	0.357	−0.009*	0.100	−0.029***	0.000	−0.041***	0.001	0.031**	0.033	0.068***	0.000	−0.017	0.125
Region3	0.016	0.392	0.067***	0.006	−0.006	0.611	−0.045**	0.038	−0.006	0.777	−0.017	0.547	−0.066***	0.000	−0.013*	0.066	−0.036**	0.011	−0.003	0.876	0.067***	0.000	−0.065***	0.000
Region4	−0.085***	0.000	0.089***	0.000	−0.005	0.675	0.050**	0.041	0.025	0.167	−0.012	0.605	−0.052***	0.000	0.010	0.253	−0.114***	0.000	0.022	0.229	0.059***	0.004	−0.019	0.414
Region5	−0.010	0.542	0.090***	0.000	−0.015	0.223	0.084***	0.001					0.068***	0.000	−0.012	0.130					−0.049*	0.061	0.034	0.224
Region6	0.027	0.125	0.019	0.413	−0.007	0.570	−0.005	0.836					0.073***	0.000	0.005	0.496					0.092***	0.000	−0.043***	0.000
Region7	−0.057***	0.000	−0.001	0.965	0.020	0.199	−0.087***	0.001													0.124***	0.000	−0.071***	0.000
Mean Outcome	0.617				0.688				0.776				0.675				0.574				0.692			
N_{total}	80369				134610				18459				1141715				111919				327181			
adj. R^2	0.459				0.399				0.178				0.419				0.521				0.300			

Notes: The table displays for labor force participation the coefficient estimates (Coef.) and the p-value of the t-test for their significance associated to equation (4) for the oldest and youngest cohort by country in a regression pooled across cohorts (i.e., the coefficients for the oldest cohorts are the interaction terms of the cohort dummy and the displayed baseline variable, the other cohort interaction terms are omitted from the output).
Source: Own calculations based on harmonized survey data (table A1).

Table A9: Labor Market Returns to Education: Wage-Employment

	Afghanistan				Bangladesh				Bhutan				India				Nepal				Pakistan				Sri Lanka			
	1960-64		1980-84		1955-59		1980-84		1960-64		1975-79		1950-54		1980-84		1950-54		1965-69		1955-59		1980-84		1950-54		1980-84	
Variable	Coef.	p-value	Coef.	p-value	Coef.	p-value	Coef.	p-value	Coef.	p-value	Coef.	p-value	Coef.	p-value	Coef.	p-value	Coef.	p-value	Coef.	p-value	Coef.	p-value	Coef.	p-value	Coef.	p-value	Coef.	p-value
(Intercept (differential Cohort FE))	0.300*	0.097	−0.044*	0.099	0.028	0.836	0.011	0.752	0.466*	0.075	0.094***	0.000	0.885***	0.000	0.082***	0.000	0.575*	0.056	0.025	0.516	0.727***	0.000	0.100***	0.000	0.870***	0.000	0.121***	0.001
Education	−0.003	0.646	−0.005	0.549	−0.034***	0.000	−0.003	0.519	0.038***	0.000	−0.032***	0.003	−0.034***	0.000	0.003	0.227	−0.036***	0.000	0.004	0.763	−0.010*	0.082	−0.020***	0.008	−0.051***	0.000	−0.043***	0.000
Education^2	0.003***	0.000	0.000	0.446	0.004***	0.000	0.000	0.874	−0.001	0.249	0.002***	0.003	0.003***	0.000	−0.001***	0.000	0.005***	0.000	−0.001	0.333	0.001***	0.000	0.001**	0.030	0.004***	0.000	0.002***	0.000
Age	−0.003	0.747			0.018***	0.003			−0.014	0.255			−0.012***	0.000			−0.013	0.329			−0.009	0.253			0.000	0.948		
Age^2	0.000	0.941			0.000***	0.001			0.000	0.242			0.000**	0.033			0.000	0.477			0.000	0.507			0.000	0.156		
Female	−0.045***	0.001	−0.014	0.445	−0.132***	0.000	0.047**	0.024	−0.126***	0.000	−0.086***	0.000	−0.045***	0.000	−0.020***	0.009	−0.129***	0.000	−0.053*	0.076	−0.034	0.153	−0.074**	0.013	−0.056***	0.000	−0.004	0.762
Urban	0.134***	0.000	−0.086***	0.002	0.155***	0.000	−0.004	0.866	0.358***	0.000	0.000	0.995	0.162***	0.000	0.013*	0.064	0.087**	0.012	0.041	0.318	0.196***	0.000	−0.078***	0.003	0.000	0.973	−0.007	0.658
Dem. Group2	−0.018*	0.063			−0.004	0.426							−0.135***	0.000			−0.018*	0.062			−0.034***	0.000			0.069***	0.000		
Dem. Group3	−0.039***	0.002											−0.214***	0.000			0.124***	0.000			−0.121***	0.000			0.355***	0.000		
Dem. Group4	0.022**	0.031											−0.222***	0.000			0.083***	0.000			0.055***	0.000						
Dem. Group5	−0.036***	0.000											−0.251***	0.000							−0.032***	0.001						
Region2	−0.100***	0.000	0.020	0.496	−0.018	0.392	0.006	0.875	−0.061***	0.001	0.029	0.274	−0.003	0.721	−0.050***	0.000	−0.003	0.924	0.029	0.431	0.191***	0.000	−0.013	0.608	−0.086***	0.000	0.039**	0.021
Region3	−0.116***	0.000	0.088***	0.006	−0.008	0.682	0.073**	0.042	0.004	0.866	0.031	0.325	0.124***	0.000	−0.103***	0.000	−0.013	0.668	0.003	0.946	0.016	0.543	−0.068**	0.040	−0.123***	0.000	0.125***	0.000
Region4	−0.095***	0.000	0.067**	0.037	−0.007	0.761	−0.034	0.381	0.013	0.491	0.005	0.863	0.117***	0.000	−0.058***	0.004	−0.076**	0.013	0.042	0.322	0.141***	0.000	−0.074*	0.053	−0.203***	0.000	0.140***	0.000
Region5	−0.085***	0.000	0.043	0.147	0.002	0.919	−0.047	0.233					0.189***	0.000	−0.093***	0.000	−0.006	0.869	−0.050	0.301					−0.166***	0.000	0.086*	0.074
Region6	−0.119***	0.000	0.065**	0.030	0.021	0.376	0.097**	0.020					0.213***	0.000	−0.071***	0.000									−0.251***	0.000	0.125***	0.000
Region7	−0.174***	0.000	0.041	0.135	0.020	0.436	−0.050	0.260																	−0.282***	0.000	0.098***	0.000
Mean Outcome	0.160				0.358				0.384				0.456				0.202				0.584				0.561			
N_{total}	42106				90977				14155				766510				8668				63453				222051			
adj. R^2	0.297				0.096				0.344				0.094				0.119				0.095				0.087			

Notes: The table displays for wage employment the coefficient estimates (Coef.) and the p-value of the t-test for their significance associated to equation (4) for the oldest and youngest cohort by country in a regression pooled across cohorts (i.e., the coefficients for the oldest cohorts are the interaction terms of the cohort dummy and the displayed baseline variable, the other cohort interaction terms are omitted from the output).
Source: Own calculations based on harmonized survey data (table A1).

Table A10: Labor Market Returns to Education: Wages

	Bangladesh				India				Pakistan				Sri Lanka			
	1955-59		1980-84		1950-54		1980-84		1960-64		1980-84		1950-54		1980-84	
Variable	Coef.	p-value	Coef.	p-value	Coef.	p-value	Coef.	p-value	Coef.	p-value	Coef.	p-value	Coef.	p-value	Coef.	p-value
(Intercept (differential Cohort FE))	3.824^{***}	0.000	1.241^{***}	0.000	3.242^{***}	0.000	0.978^{***}	0.000	1.495^{***}	0.000	0.604^{***}	0.000	4.454^{***}	0.000	−0.210	0.260
Education	0.031^{***}	0.000	-0.027^{***}	0.002	0.067^{***}	0.000	-0.117^{***}	0.000	0.008	0.286	-0.022^{**}	0.024	-0.047^{***}	0.000	0.137^{***}	0.000
Education^2	0.004^{***}	0.000	0.000	0.990	0.003^{***}	0.000	0.005^{***}	0.000	0.006^{***}	0.000	−0.001	0.340	0.009^{***}	0.000	-0.006^{***}	0.000
Age	0.047^{***}	0.000			0.022^{***}	0.003			0.089^{***}	0.000			0.038^{***}	0.000		
Age^2	0.000	0.627			0.000	0.118			-0.001^{***}	0.000			0.000	0.101		
Female	-0.165^{***}	0.006	0.083	0.197	-0.520^{***}	0.000	−0.012	0.533	-0.985^{***}	0.000	-0.243^{***}	0.000	-0.291^{***}	0.000	-0.081^{**}	0.026
Urban	0.236^{***}	0.000	-0.176^{***}	0.000	0.437^{***}	0.000	-0.160^{***}	0.000	0.197^{***}	0.000	0.007	0.845	0.128^{***}	0.000	−0.005	0.896
Dem. Group2	0.045^{***}	0.000			-0.021^{***}	0.005			-0.097^{***}	0.000			-0.086^{***}	0.000		
Dem. Group3					0.020^{***}	0.000			-0.232^{***}	0.000			0.027^{*}	0.076		
Dem. Group4					0.072^{***}	0.000			-0.100^{***}	0.000						
Dem. Group5					0.124^{***}	0.000			-0.292^{***}	0.000						
Region2	0.013	0.756	0.102^{*}	0.059	-0.357^{***}	0.000	0.049	0.170	−0.004	0.872	-0.108^{***}	0.001	-0.216^{***}	0.000	-0.259^{***}	0.000
Region3	0.043	0.250	0.092^{*}	0.056	-0.359^{***}	0.000	0.174^{***}	0.000	0.142^{***}	0.000	-0.122^{***}	0.009	-0.185^{***}	0.000	−0.060	0.255
Region4	-0.087^{**}	0.039	0.114^{**}	0.044	-0.066^{*}	0.069	0.079^{**}	0.049	0.213^{***}	0.000	-0.177^{***}	0.000	-0.126^{**}	0.021	−0.089	0.197
Region5	-0.152^{***}	0.001	0.186^{***}	0.001	-0.302^{***}	0.000	0.149^{***}	0.000					−0.033	0.671	−0.092	0.312
Region6	-0.294^{***}	0.000	0.335^{***}	0.000	-0.356^{***}	0.000	0.333^{***}	0.000					-0.190^{***}	0.000	0.069	0.182
Region7	0.133^{**}	0.010	−0.002	0.972									-0.169^{***}	0.000	-0.306^{***}	0.002
Mean Outcome	6.991				5.352				4.485				6.212			
N_{total}	31770				299798				30122				112050			
adj. R^2	0.477				0.593				0.629				0.234			

Notes: The table displays for wages the coefficient estimates (Coef.) and the p-value of the t-test for their significance associated to equation (4) for the oldest and youngest cohort by country in a regression pooled across cohorts (i.e., the coefficients for the oldest cohorts are the interaction terms of the cohort dummy and the displayed baseline variable, the other cohort interaction terms are omitted from the output).
Source: Own calculations based on harmonized survey data (table A1).

A.3 Opportunity Structure and Growth Across Types

Drawing on Peragine et al. (2014), we estimate opportunity profiles by calculating the average outcome for each type, where types are defined by shared circumstances. Comparing these profiles across birth cohorts allows us to assess how growth in outcomes—whether educational attainment, labor market participation, or consumption—affected different groups. This constitutes a type-based Opportunity Growth Incidence Curve (OGIC): how growth in an outcome has affected types differently.[38]

The OGIC can be interpreted as the rate of development of each social group in the population, where groups are defined on the basis of initial circumstances. This approach complements the Shapley decomposition presented in the main text by showing not just which circumstances matter most in aggregate, but where specific types sit in outcome distributions and how their relative positions evolved.

Figure A4 depicts opportunity profiles for years of education in the 1950-54 cohort. Types are ordered by their predicted educational attainment (x-axis), with symbols indicating urbanity (dots for urban, crosses for rural) and color indicating gender (blue for male, red for female). The figure reveals stark initial disparities, consistent with the high education IOp levels documented in section 4. Apart from Sri Lanka, educational attainment is highly concentrated: comparing the top 10% of types with the median type (P90/P50 ratios, table A7), we find ratios ranging from 1.37 in Sri Lanka to 2.44 in Afghanistan. The privileged, well-educated types are almost exclusively urban males (blue dots), while the most disadvantaged types are predominantly rural females (red crosses).[39]

Figure A1 complements this view by showing population CDFs by gender. In the 1950s cohorts, female CDFs cross 0.5 (median) well before male CDFs, confirming that females are overrepresented in the bottom half of the educational distribution.

Figure A5 presents the educational OGIC, showing changes in educational attainment by type between the 1950-54 and 1990-94 cohorts. The figure provides granular confirmation of the aggregate patterns documented earlier. While educational expansion appears progressive in absolute terms in Bangladesh, India, and Sri Lanka (positive slope of fitted spline, meaning disadvantaged types gained most), the pattern is more mixed in Bhutan, Nepal, and Pakistan, where gains appear progressive only in relative terms. Afghanistan exhibits regressive changes, with already-advantaged types gaining more education than disadvantaged types, consistent with Afghanistan's limited IOp reduction documented in figure 1.[40]

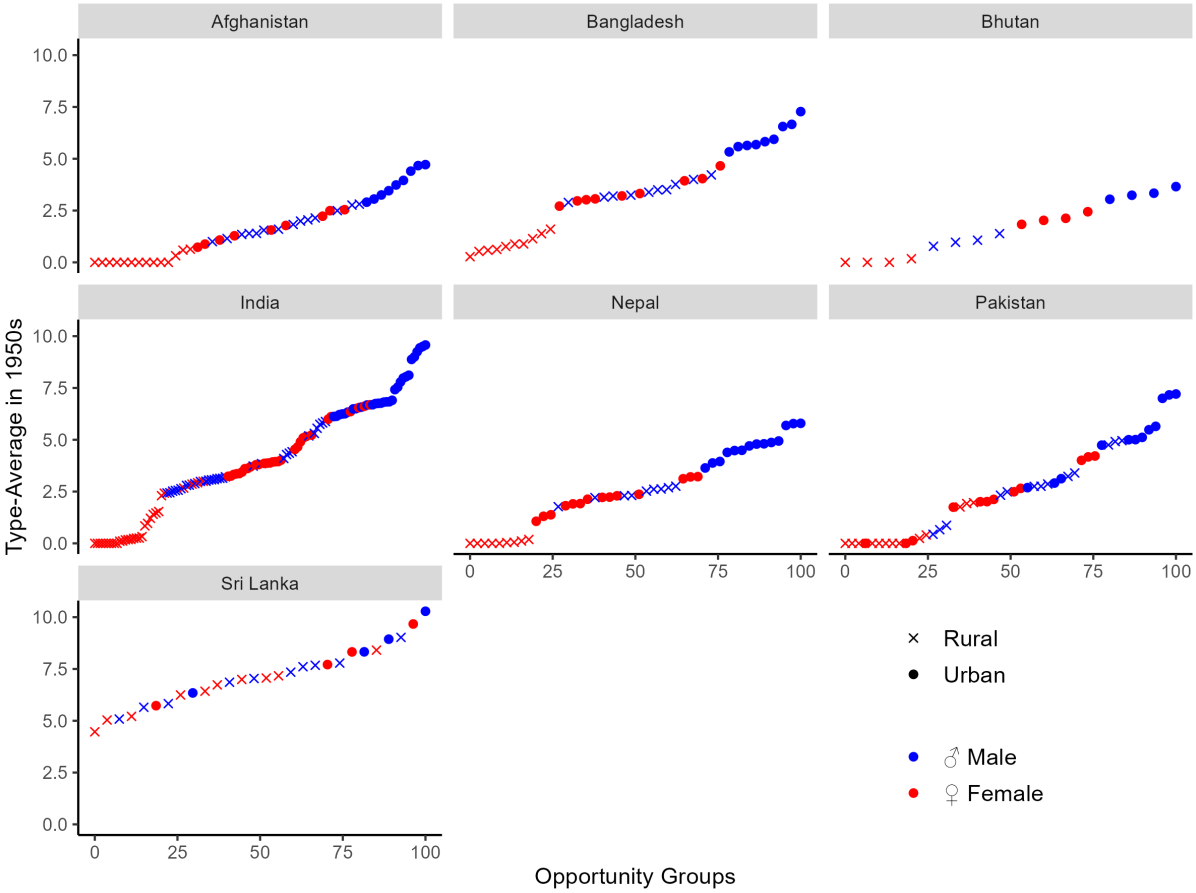
Critically, female types (red symbols) exhibit the largest gains in most countries,

[38] The OGIC translates the GIC of Ravallion and Chen (2003) into the sphere of IOp. Formally, letting $[Y_{\mu,t}] = (\mu_1(y_t), \ldots, \mu_n(y_t))$ be the distribution of type mean outcome at time (i.e., birth cohort) t, where types are ordered increasingly according to their mean, i.e.,$\mu_1(y_t) \leq \cdots \leq \mu_n(y_t)$, and $\left[\tilde{Y}_{\mu,t+1}\right] = (\tilde{\mu}_1(y_{t+1}), \ldots, \tilde{\mu}_n(y_{t+1}))$ is the distribution of type mean income at time $t+1$, where types are ordered according to their position at time t, the type OGIC is defined as $\tilde{g}(\frac{i}{n}) = \frac{\tilde{\mu}_i(y_{t+1}) - \mu_i(y_t)}{\mu_i(y_t)}, \forall i \in \{1, \ldots, n\}$. We opt for estimating the type-means $\mu_i(y_t)$ parametrically via equation (1). Non-parametric results do align qualitatively and are available on request.

[39] Sri Lanka is a notable outlier, with some female types (Sinhalese women from southern and western regions) among the educational elite and some male types (rural Sri Lanka Tamil/Moors men from the eastern and central regions) in the bottom quartile. This reflects Sri Lanka's distinct educational trajectory of early near-universal basic education (Dundar et al., 2014, 2017).

[40] Defining progressivity only in relative terms (percentage change) would be misleading given near-zero educational attainment of the most disadvantaged types in early cohorts. We therefore assess progressivity based on absolute changes in years of education.

Figure A4: Profiles 1950s - Years of Education

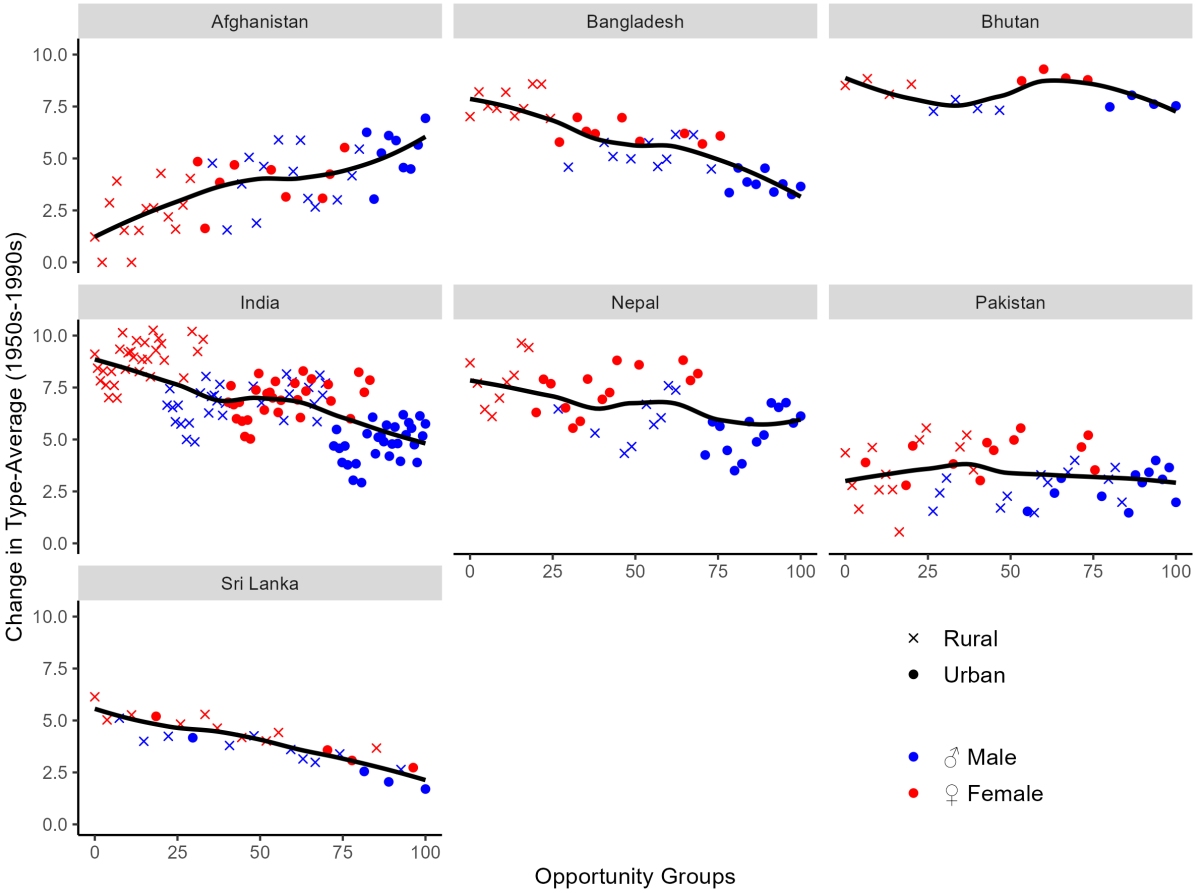


Notes: The figure depicts the opportunity profiles for years of education of the 1950-54 birth cohort across countries, i.e., the predicted average educational attainment by type (opportunity group) in the oldest cohort. Opportunity groups ordering (x-axis) is based to their predicted outcome (y-axis), i.e., the group with the lowest predicted outcome sits on position 1 whereas the group with highest predicted outcome is positioned at 100 (for comparative purposes, we rescale the rank position to 1-100 as the number of groups varies across countries due to different circumstance granularity). Hence, a group's position is simply a relative rank without an inherent scale. Group characteristics (circumstances) are indicated by the symbol (Urbanity) and its color (Gender), e.g., red crosses indicate females living in rural areas. The remaining variation in outcomes stems from the circumstances region of residence and demographic group, which are not indicated for the sake of readability. Note that opportunity groups do not necessary exhibit the same size (see figure A1).
Source: Own calculations based on harmonized survey data (table A1).

particularly among rural females (red crosses). This visually confirms gender's declining contribution to education IOp identified through the Shapley decomposition (figure 10). However, as figure A5 shows, most female types remain concentrated in the lower half of the distribution even in the 1990s cohort, and their average gains often place them at primary or lower-secondary levels—precisely the range with minimal labor market returns (section 5.2).

Figure A20 shows opportunity profiles for labor force participation in the 1950s cohort, providing a granular view of the labor market disparities identified in section 4.2. The figure reveals staggering gender-based stratification. Certain female types—such as "Muslim women in rural eastern India" and "Pashtun women in Balochistan (urban and rural), Pakistan"—face labor force participation rates consistently below 10%. Male types from the same regions exhibit participation rates of 80-95%. Conditional on labor force partic-

Figure A5: Changes in Opportunity Profiles - Years of Education



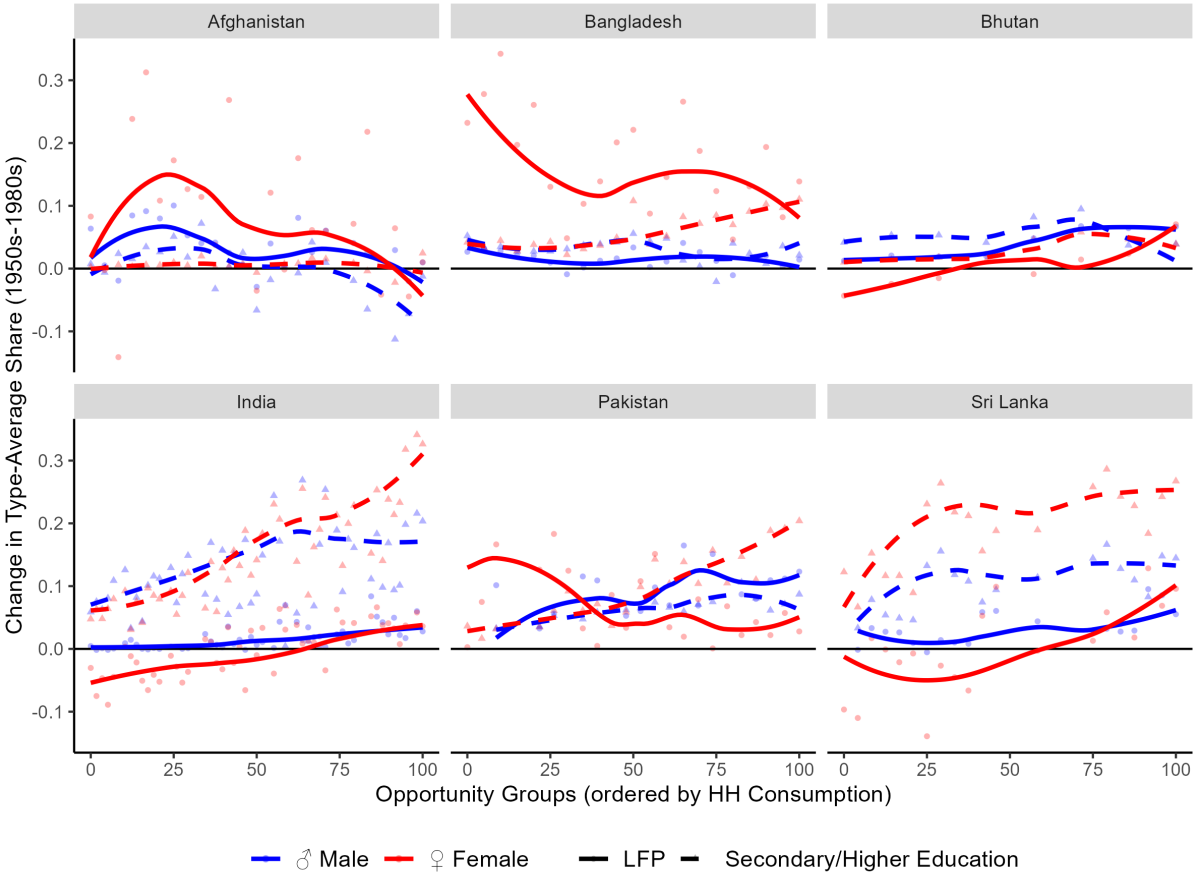
Notes: The figure depicts the change in opportunity profiles between the 1950-54 and 1990-94 birth cohorts, i.e., the difference between the predicted average educational attainment by opportunity group between the two birth cohorts. Remark, we refrain from the rescaling this change by the base cohort's educational attainment as many groups of the 1950s have zero formal education (see figure A4). Group characteristics (circumstances) are indicated by the symbol (Urbanity) and its color (Gender), e.g., red crosses indicate females living in rural areas.
Source: Own calculations based on harmonized survey data (table A1).

ipation, wage employment rates are somewhat less stratified by gender but show strong urban advantages (figure A21), consistent with higher prevalence of formal employment in urban areas and the regression results presented in section 5.2.

Figure A6 synthesizes these patterns by examining changes in both secondary/higher educational attainment (triangles, dashed lines) and labor force participation (dots, solid lines) between the 1950s and 1980s cohorts, with types ordered by their predicted consumption level.[41] Each type appears four times: male secondary education (blue triangles), female secondary education (red triangles), male LFP (blue dots), and female LFP (red dots).

[41] Cohorts are restricted by labor market data availability (maximum 1980-84 cohort). This restricts observed educational expansion primarily to basic education efforts rather than secondary expansion, especially relevant for Bhutan where secondary attainment was only 10% in the 1975-79 cohort (figure 5).

Figure A6: Changes in Opportunity: Secondary/Higher Education vs. LFP

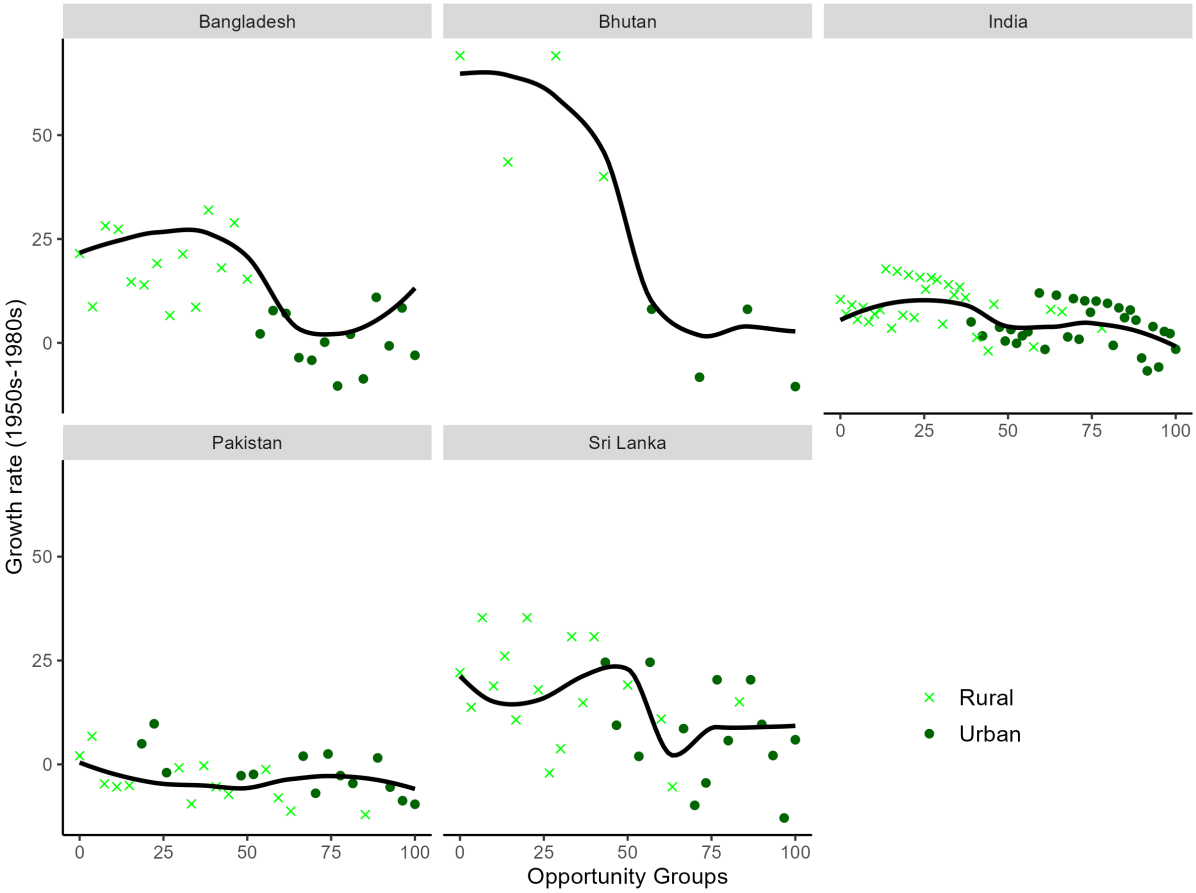


Notes: This figure depicts changes in opportunity with respect to upper secondary and higher educational attainment (triangles, dashed line) and labor force participation (dots, solid line) between the 1950s and 1980s birth cohorts by gender (blue male, red female), i.e., analogues to figure A5, the difference between the predicted average outcome (probability of having completed at least upper secondary education or probability of participating in the labor market) by opportunity group between the two birth cohorts. Opportunity groups are ordered according to household consumption per capita (for Afghanistan the male primary education attainment is used due to the absence of cohort-based consumption estimates), i.e., each opportunity group (x-axis) has 4 markers.
Source: Own calculations based on harmonized survey data (table A1).

The figure provides vivid visual confirmation of the broken transmission mechanism identified through the regression analysis and IOp decompositions. Even for males (blue), LFP gains (solid lines) generally lag behind educational gains (dashed lines), particularly in India and Sri Lanka where education improved substantially but labor force participation increased more modestly. For females (red), this disconnect is far more severe. Substantial gains in secondary educational attainment (rising red triangles, particularly in the upper half of the distribution in Bangladesh, India, Pakistan, and Sri Lanka) yielded minimal improvements in labor force participation (flat red dots). In Bangladesh, India, and Pakistan, even advantaged female types who achieved substantial secondary education gains saw negligible improvements in labor force participation rates.

This granular, type-level perspective confirms the gender-specific dimension of the puzzle established earlier: educational expansion removed gender-based barriers to schooling, driving down education IOp, but labor market participation remained blocked for women despite educational credentials, keeping labor force participation IOp high (figure 11, flat red dots in figure A6).

Figure A7: Growth Incidence (OGIC) - Consumption



Notes: The figure depicts the growth in average per capita consumption by opportunity groups (types) between the 1950s and 1980s birth cohort across countries. To facilitate the distributional comparison across countries, consumption levels are rescaled by the country-specific 1950s cohort median consumption level (see footnote 42). Note for India, the 1970-74 cohort serves as comparison to the 1950s given a change in survey methodology in later surveys (see section A.1).
Source: Own calculations based on harmonized survey data (table A1).

Figure A7 visualizes consumption OGICs by plotting each type's consumption growth (1950s to 1980s cohorts) against the type's rank in the 1950s consumption distribution. Color coding distinguishes rural (green crosses) from urban (dark green dots) types.[42]

[42]Consumption per capita as welfare aggregate is standardized by the World Bank (2024c). Yet, the exact bundle of consumption items, the recall period of expenditure and methodological choices tend to vary across countries such that directly comparing the monetary value of this welfare aggregate can be misleading. Hence, we standardize consumption by dividing the monetary, CPI-adjusted value by the median consumption in the oldest (i.e., 1950s) cohort for a given country. This renders an interpretable y-axis for the comparison of opportunity profiles across country, i.e., an opportunity ratio which regards the opportunity group's mean outcome relative to the median outcome in the 1950s. Fixing the denominator the 1950s cohort allows us to capture the welfare gains associated to economic growth across the period of survey collection, i.e., how does welfare compare during prime working age across cohorts.

For example, being aged 35-54, the most disadvantaged opportunity group of Bhutan, rural households in the East, has an average consumption per capita of 0.77 with respect to the cohort's median of 1820 2010-Nu for those born in the 1950s, whereas individuals of this group born in the early 1980s consume per capita 2650 2010-Nu, i.e., 1.45 of the 1950s median consumption. Hence, the groups change in the consumption opportunity ratio across those 30 year birth cohorts, 1.45-0.77=0.68, reflects well the large catch-up process of rural population in Bhutan over the last two decades (see table A12 for the robustness of IOp with respect to life cycle bias).

The bottom half of types are nearly exclusively rural in all countries except Pakistan, where urban types from Balochistan—reflecting that province's limited economic development—appear among the most disadvantaged.

Bangladesh and Bhutan exhibit progressive consumption growth—rural types (green) experienced higher growth than urban types (black)—consistent with those countries' modest consumption IOp improvements (figure 3). Rural catch-up processes and increasing urbanization both contribute to this pattern.[43] India, Pakistan, and Sri Lanka show flatter or regressive patterns, consistent with stagnant consumption IOp in those countries.

These consumption OGIC results align with the labor market findings. Countries where some disadvantaged types—particularly rural types—gained labor market access (as evidenced by shifting opportunity profiles) also exhibit progressive consumption growth and modest consumption IOp improvements. Where labor market access remained severely restricted for disadvantaged types, consumption opportunities stagnated despite educational expansion.

[43]Data restrictions prevent studying urbanization's direct impact, as urbanity status is only known at data collection, not birth (see section 3). However, examining residential patterns via population CDFs (figure A12) and population shares changes (figure A13), we find younger cohorts more likely to live in urban areas. This likely understates true urbanization at birth/childhood—when educational opportunities are shaped—but applies to both education and consumption. Note that Bhutan experienced extremely rapid urban growth (an additional 25% of population in urban areas since 1990; World Bank (2025)), concentrated in the West zone driven by capital Thimphu.

A.4 Importance of Parental Background

The main analysis omits parental background - a major contributor to unequal opportunities (Brunori et al., 2013; Singh, 2012). Given the severe data limitations described in section 3, i.e., the absence retrospective questions on parental education, we turn individuals coresiding with their parents in order to extract information on parental background. This introduces two kinds of distortions in comparison to the previous analysis based on individuals with completed educational careers: (i) truncation, i.e., educational attainment has to be top-coded as post-secondary/tertiary education completion cannot be observed; (ii) co-residency bias (Emran et al., 2018), i.e., individuals co-residing with the parents are not representative for the general population after a certain age.

Comparing the coresident individuals and the full population with completed educational careers, we find large difference on the amount educational inequality between the two samples, i.e., we capture 38 to 86 percent less total inequality (column 1 table A11). However, conditional on our limited age sampling frame (15-18 years), the coresident population is largely representative for the given age group (see column 3 table A11 and figure A8). Therefore, the distortion in total outcome inequality between full and coresident sample mainly arises from truncation (i.e., the difference educational inequality due the limited age of the sampling frame, see column 2). This distortion also translates into IOp estimates, i.e., considering the same limited set of circumstance, estimated IOp is up to 75 percent lower for the coresident sample (column 4).[44]

Naively comparing the IOp estimates of the full population with the limited set of circumstances to the IOp estimates of the coresident sample with the extended set of circumstances (i.e., additionally including parental education), we do not observe higher IOp for the coresident sample due to the truncation distortion (column 5). Yet, we obtain a proxy for the importance of parental background when comparing the IOp estimates of the coresident sample with and without it (column 6).[45]

We can validate quality of this proxy using on a limited set of surveys which include a direct parental background question, i.e., we can compute the difference in IOp due to the inclusion of parental education for the full population of a birth cohort without the age restrictions of the coresident analysis (including all secondary and tertiary educational attainment) and compare it with the proxy (column 8 vs. 5). Further, we can check whether the limited co-resident distortion found with respect to total inequality (column 3) also generalizes with respect to IOp by truncating the outcome distribution in these surveys which direct parental background questions (column 7). For both, overall performance and coresident distortion in IOp estimates, we find varying results by country and birth cohorts which cannot properly generalize due to the small number of surveys with direct parental background questions and their timing with respect to the other surveys available for usage in the coresident analysis.

[44] The only exception is Sri Lanka which exhibits a negative co-residency bias, i.e., those individuals coresiding with their parents a slightly less educated than their full age group. This negative bias tends increase across birth cohorts which could be reconciled by increasing importance of secondary/tertiary education which may require to move-out.

[45] Adding circumstances (e.g., parental education) tends to increase IOp as it allows for a finer partition of types and, hence, mechanically accounts for more inequality in the outcome distribution (Fleurbaey et al., 2017).

Table A11: Years of Education Coresident Analysis - Distortion vs. Proxy

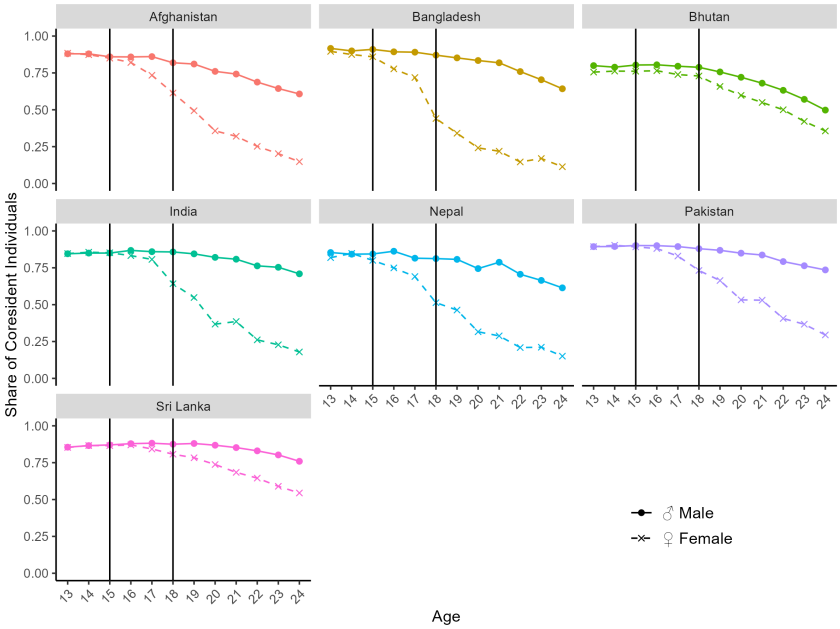
	Total Inequality			Relative IOp				
Country & Cohort	Cores-Full	Age-Full	Cores-Age	limited C	naive	proxy	trunc	true
Afghanistan 1990-1993	-81.00	-81.90	5.10	-59.40	-58.90	1.20		
Afghanistan 1994-1997	-80.90	-81.90	5.50	-72.40	-68.00	15.90		
Afghanistan 1996-1999	-77.40	-78.20	3.70	-67.20	-64.90	6.90		
Afghanistan 1999-2002	-76.70	-77.30	3.00	-65.00	-62.20	7.90		
Afghanistan 2002-2005	-78.90	-79.30	2.00	-48.40	-48.10	0.60		
Bangladesh 1982-1985	-69.10	-69.40	1.00	-47.90	-20.60	52.40		
Bangladesh 1987-1990	-63.60	-63.40	-0.60	-54.40	39.60	206.50		
Bangladesh 1992-1995	-69.10	-68.40	-2.30	-14.10	78.10	107.30		
Bangladesh 1998-2001	-63.90	-63.10	-2.30	109.70	229.80	57.30		
Bangladesh 2004-2007	-69.00	-68.50	-1.80	145.10	264.40	48.70		
Bhutan 1985-1988	-70.60	-71.10	1.70	-57.40	-55.20	5.20	3.60	
Bhutan 1994-1997	-69.60	-68.20	-4.40	-10.80	5.40	18.00	5.60	
Bhutan 1999-2002	-80.50	-80.50	-0.20	18.80	36.30	14.70		
Bhutan 2004-2007	-87.00	-85.30	-12.10	34.40	116.40	61.10		
India 1975-1978	-64.50	-65.40	2.70	-76.90	-64.60	53.30	75.90	20.50
India 1986-1989	-69.60	-67.60	-6.40	-29.80	-8.30	30.80	52.40	43.00
India 1987-1990	-70.00	-68.50	-4.90	-28.40	-11.10	24.10	42.30	42.80
India 1991-1994	-70.80	-70.90	0.50	-3.70	32.10	37.20	13.90	
India 1993-1996	-66.90	-67.20	1.00	18.60	54.80	30.50	6.50	
India 2004-2007	-91.10	-90.90	-3.10	67.60	121.10	31.90	8.50	
Nepal 1977-1980	-62.00	-63.60	4.20	-62.60	-62.30	0.70	-0.00	7.70
Nepal 1985-1988	-65.30	-65.10	-0.50	-51.50	-40.00	23.60	10.80	12.70
Nepal 1992-1995	-63.00	-62.70	-0.90	-36.00	-17.90	28.20	29.60	
Nepal 1993-1996	-63.00	-62.80	-0.40	-36.60	-19.80	26.50	27.20	
Pakistan 1973-1976	-75.20	-76.50	5.60	-71.60	-68.30	11.80	1.00	
Pakistan 1989-1992	-74.00	-74.40	1.80	-59.60	-47.70	29.30	28.00	
Pakistan 1992-1995	-72.50	-72.90	1.40	-59.70	-43.60	40.00		
Pakistan 1993-1996	-72.30	-72.70	1.40	-59.20	-41.70	42.80		
Pakistan 1994-1997	-71.70	-72.20	1.60	-58.40	-41.80	40.10		
Pakistan 1995-1998	-71.20	-71.60	1.30	-56.30	-39.80	37.70		
Pakistan 1996-1999	-71.50	-71.90	1.30	-56.80	-39.90	39.10		
Pakistan 1997-2000	-71.30	-71.70	1.30	-54.70	-38.30	36.30		
Pakistan 2000-2003	-72.20	-72.40	0.60	-59.10	-38.00	51.80		
Pakistan 2001-2004	-70.40	-70.60	0.70	-62.60	-47.50	40.60		
Sri Lanka 1972-1975	-39.20	-36.00	-4.90	-18.50	50.30	84.50		
Sri Lanka 1977-1980	-60.30	-56.70	-8.40	37.00	108.10	51.90		
Sri Lanka 1984-1987	-77.30	-75.50	-7.30	58.00	132.50	47.20		
Sri Lanka 1988-1991	-82.20	-79.80	-11.90	99.20	178.80	40.00		
Sri Lanka 1991-1994	-80.80	-78.90	-9.30	80.60	144.80	35.60		
Sri Lanka 1994-1997	-87.70	-86.80	-6.70	81.20	135.80	30.10		
Sri Lanka 1998-2001	-93.70	-92.90	-10.90	152.70	247.30	37.40		
Sri Lanka 2001-2004	-92.40	-92.00	-5.90	221.70	350.90	40.10		

Notes: The table displays in % changes of estimate displayed in column header for: (i) restricting the full sample to co-residing individuals in terms of **total inequality** (1) and relative IOp with **limited C** (3) whereas (2) restricts the sample of individuals aged 15-18 to those co-residing in terms of **total inequality**; (ii) **naive** being the difference between full sample with limited C and coresident sample with extended C; (iii) **proxy** being the difference between limited and extended C for coresident sample; and (iv) **trunc** and (v) **true** being the difference between limited and extended C for full sample (only surveys with direct parental background question) with truncated (i.e., top-coded) and full outcome, respectively.
Source: Own calculations based on harmonized survey data (table A1).

Given that inferring the exact differential of the quantity of inequality explained by circumstances to be attributed to parental background from the proxy appears not entirely credible, we turn to the relative circumstance importance of parental background in the cohort-based estimates in order to trace out potential changes (figure A9). As indicated by the sizable increases in IOp estimates due to parental background inclusion, we find that parental education accounts for around 50% of educational IOp in all countries for the most recent cohorts except for Afghanistan and Bhutan. Regarding trends

across birth cohorts,[46] a heterogeneous picture is found. Bhutan, Nepal and Pakistan exhibit a substantial increase in the importance of parental background between 1970s and 1990s birth cohorts which could be explained by the increasing importance of secondary education, i.e., while initially location and gender where barriers to any kind of education, primary education became widely available such that educational inequality arise at subsequent levels of educational attainment for younger cohorts. For India and Sri Lanka, the large importance of parental background for the cohorts born in the 1970s (57% and 67%) reduces across birth cohort but remains around 50% for individuals born in the late 1990s. An interesting and encouraging development is the continued reduction across birth cohorts since the early 1990s after the large increase since the 1970s.

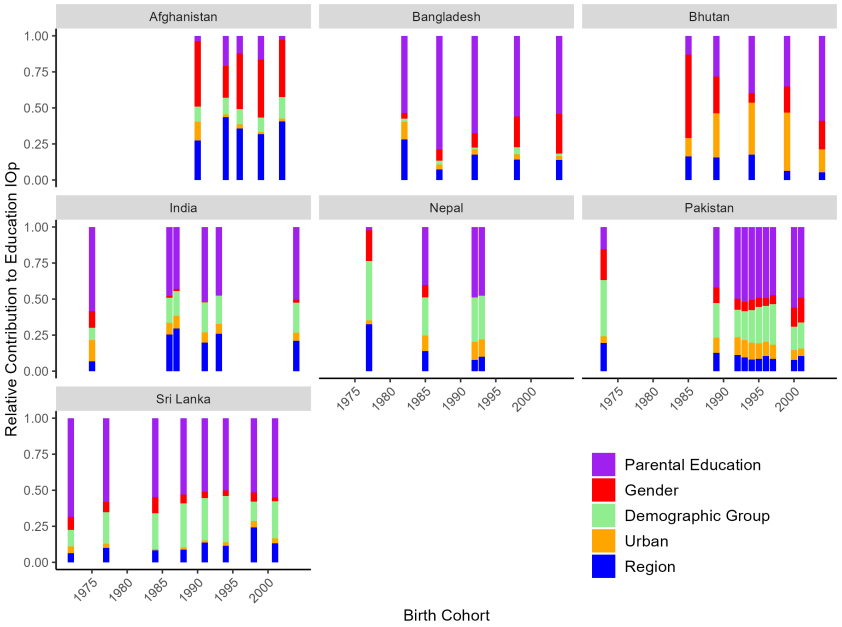
Figure A8: Sampling Frame: Coresident Share by Gender



Notes: The figure depicts the share of coresident individuals by age and gender for each country. The sampling frame (i.e., age 15-18) is marked with black lines. We observe that female individuals move out well before male ones. Given that we want to estimate IOp within the full population, we rely on a uniform sampling frame and top-coding across gender which trades-off the representativeness of the sample with the extend of educational inequality captured. Coresident shares are below 1 due to the presence of multi-generation household as individuals who are grand-children of the household head are not classified as coresident (i.e., only individuals being the child of the household head are registered as coresident). **Source**: Own calculations based on harmonized survey data (table A1).

[46] In the presence of changing moving-out patterns across birth cohorts, the trends in the importance of parental background across coresident cohorts are compound the potential changes in parental background importance and changes in moving-out patterns.

Figure A9: Circumstance Importance: Years of Education - Coresident Sample

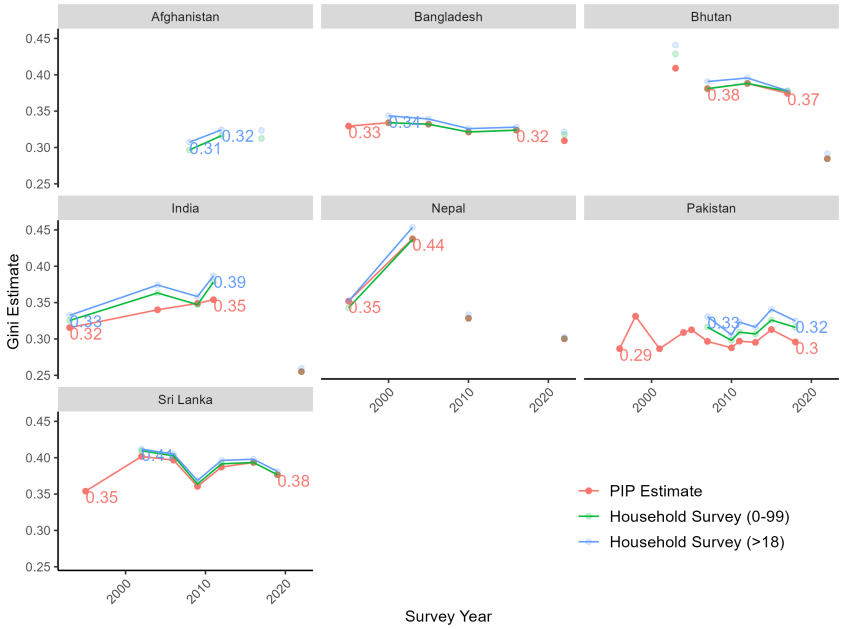


Notes: The figure depicts the relative importance of the different circumstances for IOp in years of education as obtained via a Shapley decomposition (footnote 30).

A.5 Robustness

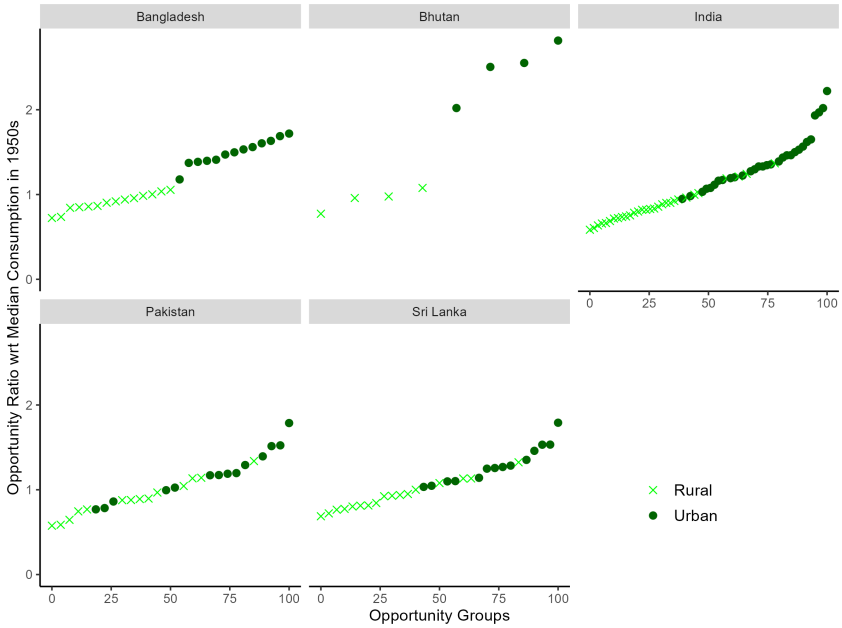
A.5.1 Robustness: Consumption

Figure A10: Total Inequality Comparison PIP - Consumption



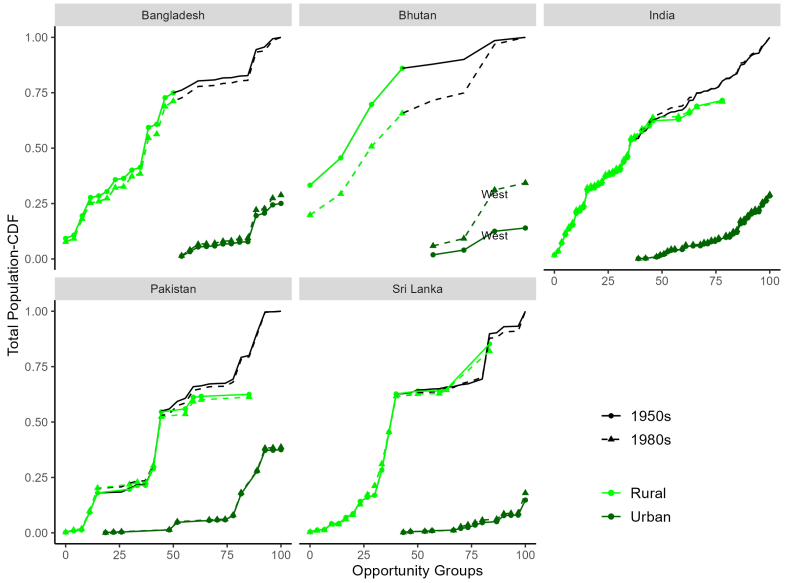
Notes: The figure compares our cross-sectional estimates of total inequality for different sample definitions as measured by the Gini coefficient with the World Bank's Poverty and Inequality Platform (PIP). Following the Global Estimates of Opportunity and Mobility (GEOM) database, our cross-sectional estimates presented in table 2 are based on individuals aged at least 18 years (blue). Slight deviations from the PIP Gini estimates (red) can be mainly be reconciled by extending the sample to the full population without age restriction (green). The absence of a connecting line between points indicates the non-comparably of the consumption estimates across the concerned cross-sections. Note, for India a major revision of the PIP estimates has been released in September 2025 due to a new harmonization between NSS and HCES (see section A.1). Given that our SARMD data has not been updated accordingly, PIP estimates of the previous release (September 2024) are displayed for India apart from the 2022 estimate.

Figure A11: Profiles 1950s - Consumption



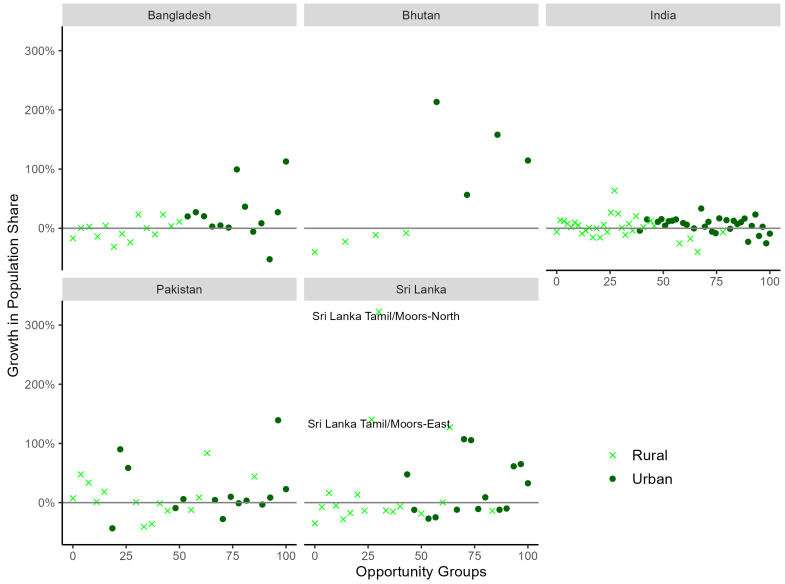
Notes: The figure depicts opportunity profiles for per capita consumption of the 1950s birth cohort across countries, i.e., the predicted average consumption level by type (opportunity group) in the oldest cohort. To facilitate the distributional comparison across countries, consumption levels are rescaled by the country-specific median (see footnote 42).
Source: Own calculations based on harmonized survey data (table A1).

Figure A12: Population CDFs - Consumption: Changes in Urbanity



Notes: The figure depicts the Cumulative Distribution Function (CDF) of the rural (green) and urban (orange) subpopulations with respect to the total population across countries and opportunity groups in the 1950s (solid) and 1980s (dotted) cohorts, i.e., the proportion of the total population with an predicted consumption level smaller than or equal to opportunity group indicated on the x-axis associated to urban/rural areas. The triangle/dotted line is always below (above) the dots/solid line for rural (urban), i.e., a lower share of the population of the younger cohorts is living in rural areas compared to the older cohorts. However, this underestimates the urbanization tendency across cohorts as the urbanity status is only recorded at the age of data collection (i.e., old cohorts are measured at later age (age 45-54) and, hence, at more advanced stages of the urbanization process than if they would have interviewed at the same age as the younger cohorts (age 35-44)). Note, Bhutan has experienced extremely rapid urban growth (additional 25% percent of the population in urban areas since 1990, World Bank, 2025)), particularly concentrated in the West zone driven by its capital Thimphu.
Source: Own calculations based on harmonized survey data (table A1).

Figure A13: Growth in Population Shares - Consumption: Changes in Urbanity



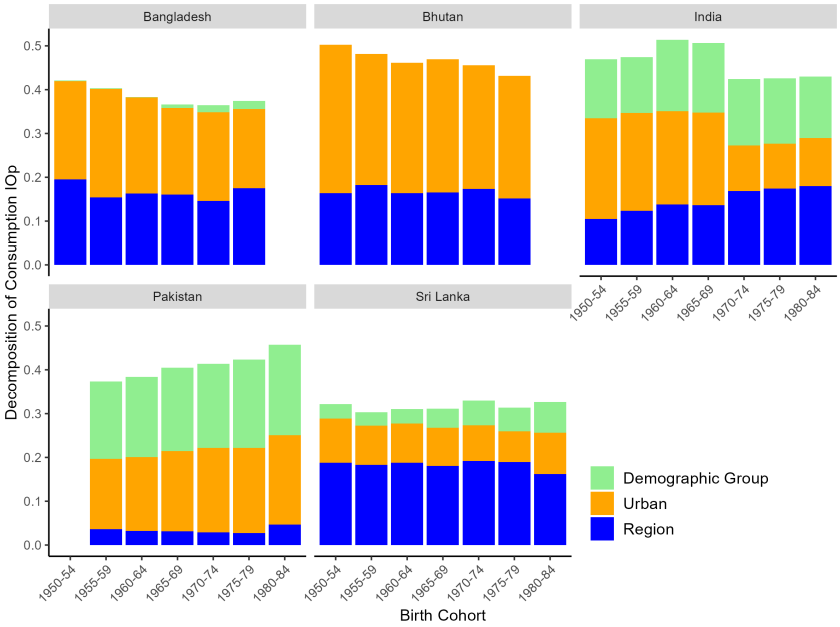
Notes: The figure depicts the growth in the population shares of the opportunity groups between the 1950s and 1980s birth cohort across countries, i.e., for a given type, the difference between its weighted share of the total 1950s and 1980s cohort population divided by its weighted share of the total 1950s cohort population. Apart from the strong urbanization in Bhutan (more than doubling of the population shares for 3 of the 4 urban types) already noted in figure A12, the drastic increase in the rural population of the Northern and Eastern provinces in Sri Lanka (Sri Lanka Tamil and Moors are the largest demographic groups in those provinces, $> 80\%$) can be linked to end of civil conflict in 2009 which throughout the 1990s and 2000s had caused high levels of displacement and forced migration.
Source: Own calculations based on harmonized survey data (table A1).

Table A12: Cohort-Age Group Table: Bangladesh

Age	Birth Cohort						Avr. IOp /
Group	1950-54	1955-59	1960-64	1965-69	1970-74	1975-79	Total Obs.
35-39			0.428	0.419	0.359	0.399	0.401
			1307	3692	3792	7096	15887
40-44		0.489	0.396	0.359	0.367	0.337	0.39
		858	2989	2997	6188	5275	18307
45-49	0.455	0.417	0.38	0.363	0.37		0.397
	688	2612	2915	5393	5725		17333
50-54	0.41	0.363	0.385	0.355			0.378
	2008	2108	4376	4044			12536
Avr. IOp	0.433	0.423	0.397	0.374	0.365	0.368	0.392
Est. IOp	0.421	0.403	0.383	0.366	0.364	0.374	0.385
Total Obs.	2696	5578	11587	16126	15705	12371	64063

Notes: The table depicts cohort-based estimates of consumption IOp and the the underlying sample size (N) at different ages (5 year age groups) of a respective birth cohort, i.e., column indicate the birth cohort and rows the age groups. The bottom of the table displays the average IOp across the age groups of a given cohort (Avr. IOp) as well as the IOp estimate for the cohort based on all its observations (Est. IOp).
Source: Own calculations based on harmonized survey data (table A1).

Figure A14: Decomposing Consumption IOp by Circumstances

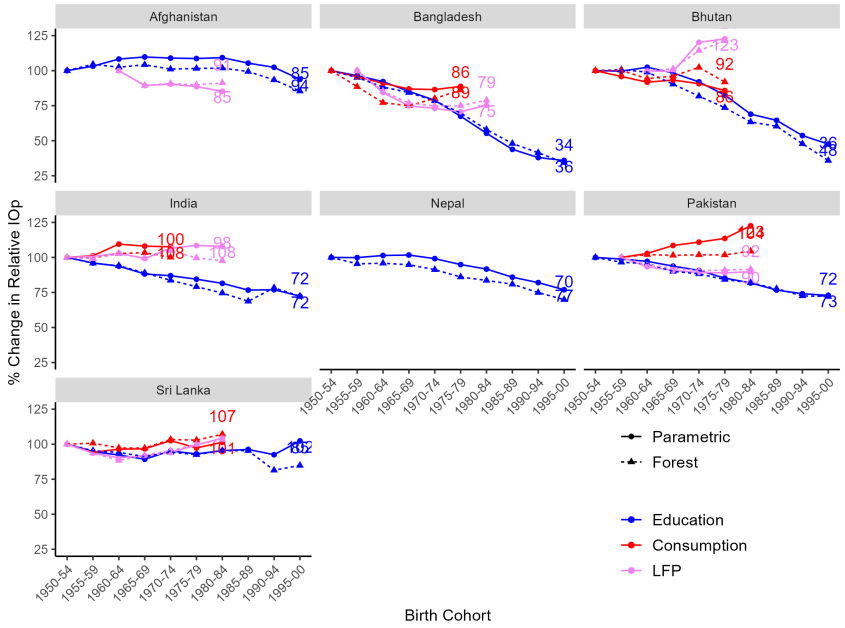


Notes: The figure depicts the importance of the different circumstances for consumption IOp as obtained via a Shapley decomposition (see footnote 30). While the hight of the columns represents the extend of relative IOp, its colored segments indicate the contribution of the corresponding circumstance to it. No systematic substantial changes in the composition of IOp can be observed.
Source: Own calculations based on harmonized survey data (table A1).

A.5.2 Robustness: Estimation Technique

The existing literature has showed that the results of the IOp measurement exercise to be potentially sensitive to the empirical model adopted. In particular, two of the major shortcomings the employed standard linear parametric estimation are that (i) they are based on a pre-specified number of circumstances and (ii) depend on the researcher's choice of including interactions of the selected circumstances into the counterfactual estimation. While excluding potential circumstance or pre-aggregating circumstance variables may cause underestimated IOp, including all possible circumstance interactions may lead to an overestimated IOp.[47] To address the latter, Brunori et al. (2023) introduced a novel approach of analyzing IOp by using regression trees that let the algorithm choose the most relevant circumstances in a statistically significant way from the set of circumstances. Averaging over a collection of trees, i.e., estimating a forests allows to alleviate the issue of sampling sensitivity of a single tree. Hence, we use regression forests to validate the results of the standard parametric analysis.

Figure A15: Evolution of IOp: Parametric vs. Forest Estimation



Notes: The figures depicts, analogous to figure 3, the relative evolution of IOp across countries, cohorts and outcome dimensions based on two estimation methods (parametric/forest), i.e., setting to 100 the IOp of the earliest birth cohort for which all outcomes are available for each country, the index represents the change in percentage points with respect to the level of IOp of this cohort. Apart from the consumption IOp evolution in Pakistan, we note the overall alignment of estimates across the two estimation methods for both outcomes.
Source: Own calculations based on harmonized survey data (table A1) but without the inclusion of the latest survey waves for the forest-based estimates of education due computation time limitations.

[47] The resulting trade-off between the two approaches has to be balanced given the data availability, i.e. a linear specification might be too restrictive, whereas including the full set of circumstances' interactions might cause very large sampling variance of the estimated counterfactual distribution when the number of observations per type is limited (see Brunori et al., 2019b).

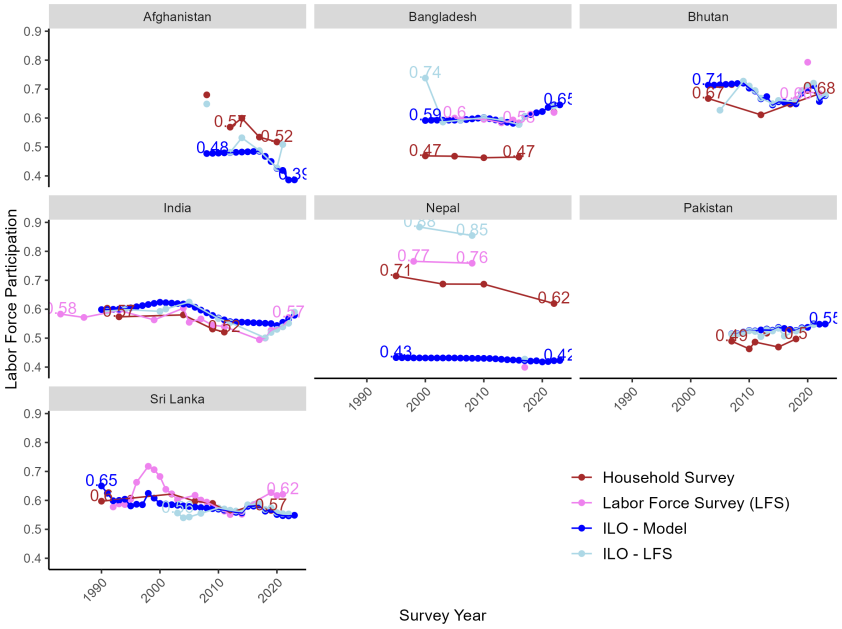
A.5.3 Results & Robustness: Labor Market

Table A13: Survey-specific Labor Market Outcome Estimates for India

	LFP				Wage-Employment				Wages			
	Share		IOp		Share		IOp		Gini		IOp	
	EUS	PLFS	EUS	PLFS	EUS	PLFS	EUS	PLFS	EUS	PLFS	EUS	PLFS
1950-54	0.68		0.20		0.45		0.13		0.57		0.61	
1955-59	0.68		0.20		0.44		0.13		0.58		0.63	
1960-64	0.68		0.21		0.45		0.12		0.57		0.64	
1965-69	0.68	0.64	0.20	0.24	0.44	0.44	0.13	0.14	0.54	0.52	0.60	0.55
1970-74	0.67	0.67	0.22	0.22	0.47	0.44	0.12	0.13	0.52	0.49	0.58	0.53
1975-79		0.67		0.22		0.46		0.12		0.46		0.54
1980-84		0.66		0.22		0.48		0.11		0.44		0.54

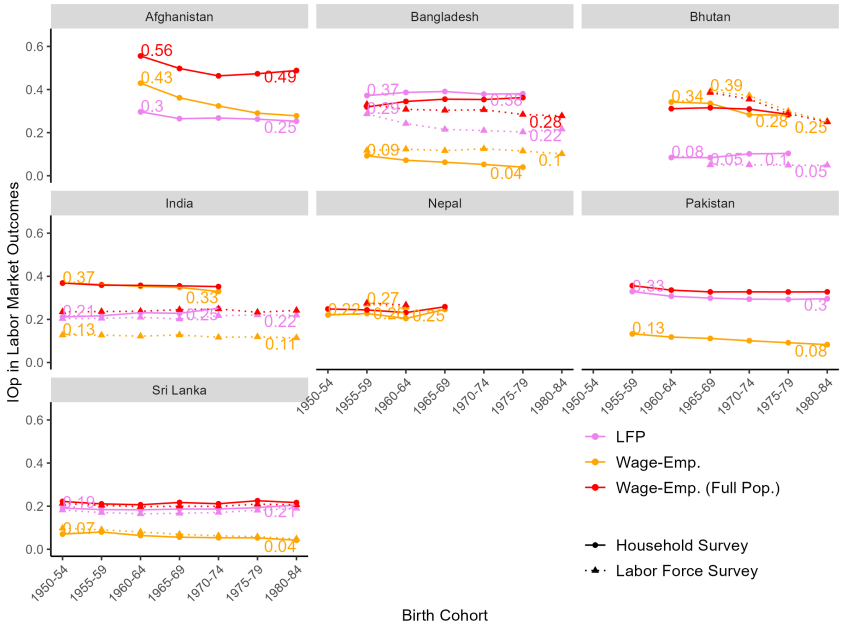
Notes: The table reports the estimates for India by surveys (EUS 1983-2011, PLFS 2017-2022) across cohorts for which a pseudo-panel can be constructed (i.e., at least two cross-sections for a given cohort). All cohort-based estimates presented throughout the paper use the PLFS only for the youngest cohorts (1975-79, 1980-84).
Source: Own calculations based on harmonized survey data (table A1).

Figure A16: LFP Estimates across different Data Sources (age 15-64)



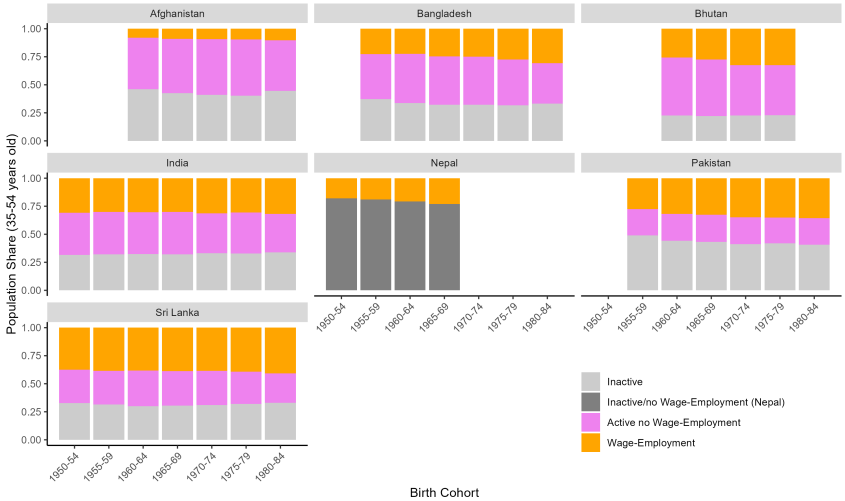
Notes: The figures illustrates the difficulty for a common definition of LFP across countries, years and survey waves. While ILO model-based LFP estimates (blue) facilitate standardized comparisons across countries and over time, their labor-force survey based (ligthblue) counterpart exhibit substantial volatility in line with changes in national definitions/methodologies. Hence, the ILO model-based estimates may differ from national estimates (see annex table A2.1.1 of World Bank (2024b) for further discussion).

Figure A17: IOp in LFP & Wage-Employment by Data Source & Sample Definition



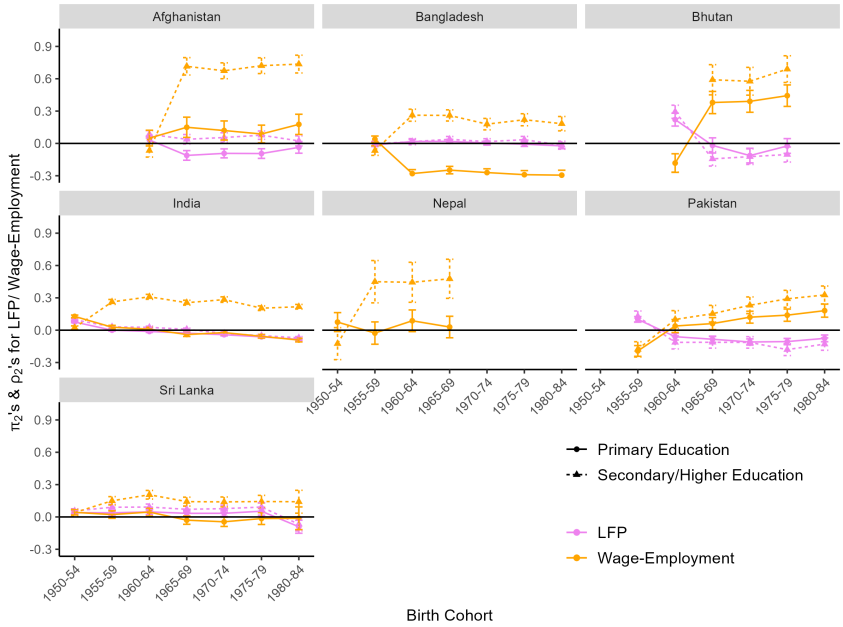
Notes: The figure depicts across birth cohorts IOp in LFP (brown) and Wage-Employment. For the latter, two different sample populations (i.e., denominators) are considered: the active midlife working age population (age 35-54) as in the main analysis (orange) and the full midlife working age population (red). For all countries and surveys (with the exception of the Bangladesh's household survey which exhibits unreliable wage-employment reporting), the red lines are above the brown ones, indicating that IOp is higher for finding wage employment than for participating in the labor market. See section A.1 for details.
Source: Own calculations based on harmonized survey data (table A1).

Figure A18: Evolution Composition prime working age population (35-54 years)



Notes: The figure depicts across birth cohorts the population composition with respect to the labor market status: inactive (gray), active but without wage employment (violet) or wage employed (orange). For Nepal, we cannot credibility distinguish inactive and active but without wage employment given overreported labor force participation (see section A.1).
Source: Own calculations based on harmonized survey data (table A1).

Figure A19: Urban Premia in Educational Returns: Primary vs. Upper Secondary



Notes: The figures depicts the predicted differential change in the probability of participating in the labor market (violet) and being a wage employee (orange, conditional on participating) for primary education (solid line) and secondary or higher education (dotted line) for urban dwellers, i.e., the coefficients π_2 (primary) and ρ_2 (secondary) of equation (3).
Source: Own calculations based on harmonized survey data (table A1).

Table A14: Labor Market Returns & Differentials: LFP

	Afghanistan				Bangladesh				Bhutan				India				Pakistan				Sri Lanka			
	1960-64		1980-84		1955-59		1980-84		1960-64		1975-79		1950-54		1980-84		1955-59		1980-84		1950-54		1980-84	
Variable	Coef.	p-value	Coef.	p-value	Coef.	p-value	Coef.	p-value	Coef.	p-value	Coef.	p-value	Coef.	p-value	Coef.	p-value	Coef.	p-value	Coef.	p-value	Coef.	p-value	Coef.	p-value
(Intercept (differential Cohort FE))	0.356***	0.007	−0.190***	0.000	−0.358***	0.000	0.175***	0.000	0.363	0.138	−0.001	0.968	0.415***	0.000	−0.017*	0.098	0.297***	0.003	0.050	0.129	0.249***	0.000	−0.262***	0.000
Age	0.037***	0.000			0.064***	0.000			0.033***	0.004			0.031***	0.000			0.032***	0.000			0.030***	0.000		
Age^2	0.000***	0.000			−0.001***	0.000			0.000***	0.001			0.000***	0.000			0.000***	0.000			0.000***	0.000		
Primary	−0.096	0.195	0.105	0.284	0.044	0.101	−0.211***	0.000	−0.015	0.814	−0.038	0.583	0.001	0.910	0.013	0.391	−0.050	0.160	0.042	0.328	0.063***	0.000	0.303***	0.000
Secondary	0.048	0.561	−0.015	0.891	0.034	0.476	−0.078	0.338	−0.022	0.682	0.001	0.983	0.003	0.840	0.020	0.320	−0.062	0.424	0.059	0.483	0.091***	0.000	0.276***	0.000
Female	−0.661***	0.000	0.035***	0.008	−0.709***	0.000	0.220***	0.000	−0.256***	0.000	−0.056***	0.004	−0.522***	0.000	−0.017***	0.002	−0.652***	0.000	0.017	0.287	−0.409***	0.000	−0.014	0.678
Urban	−0.050***	0.001	−0.023	0.229	−0.012	0.297	0.014	0.510	−0.240***	0.000	0.017	0.506	−0.131***	0.000	0.062***	0.000	−0.106***	0.000	0.014	0.473	−0.082***	0.000	0.195***	0.001
Dem. Group2	−0.020	0.421	0.110***	0.002	−0.021	0.131	−0.097***	0.002					0.059***	0.000	0.030***	0.009	0.033	0.210	−0.030	0.390	0.002	0.875	0.103**	0.020
Dem. Group3	0.131***	0.000	0.171***	0.002									−0.011*	0.094	0.000	0.970	0.101***	0.000	0.039	0.213	0.108***	0.000	0.031	0.585
Dem. Group4	0.178***	0.000	−0.162**	0.014									−0.106***	0.000	0.012	0.272	−0.107***	0.001	−0.002	0.951				
Dem. Group5	−0.027	0.156	0.077***	0.005									−0.063***	0.000	0.008	0.410	0.095***	0.001	0.040	0.311				
Region2	−0.034*	0.075	0.182***	0.000	0.014	0.493	−0.086**	0.036	0.042*	0.054	0.018	0.528	−0.008	0.234	−0.039***	0.000	−0.081***	0.005	0.025	0.516	0.082***	0.000	0.222***	0.001
Region3	−0.015	0.594	0.178***	0.000	0.017	0.356	−0.127***	0.000	−0.010	0.686	−0.035	0.290	−0.084***	0.000	−0.015	0.146	−0.039	0.209	−0.059	0.167	0.106***	0.000	0.125*	0.089
Region4	−0.100***	0.000	0.107***	0.000	0.002	0.926	0.000	0.995	0.022	0.252	−0.020	0.454	−0.074***	0.000	0.015	0.292	−0.143***	0.000	−0.008	0.815	0.000	0.994	−0.015	0.866
Region5	−0.019	0.351	0.095***	0.000	0.008	0.672	0.015	0.712					0.091***	0.000	−0.018	0.150					−0.112***	0.005	0.102	0.185
Region6	0.011	0.655	0.079**	0.020	0.011	0.589	−0.074*	0.060					0.088***	0.000	0.029***	0.007					0.094***	0.000	0.149**	0.036
Region7	−0.096***	0.000	0.086***	0.007	0.022	0.368	−0.127***	0.003													0.113***	0.000	0.135	0.107
Primary x Female	−0.186***	0.000	0.168**	0.015	−0.059***	0.000	−0.016	0.507	−0.164**	0.019	0.073	0.342	−0.219***	0.000	0.123***	0.000	−0.172***	0.000	−0.001	0.975	−0.144***	0.000	−0.017	0.636
Secondary x Female	0.347***	0.000	−0.126	0.208	0.093**	0.044	−0.152***	0.009	−0.036	0.693	0.203**	0.042	−0.082***	0.000	0.012	0.438	0.010	0.850	−0.095*	0.090	0.115***	0.000	−0.113***	0.003
Primary x Urban	0.032	0.359	0.059	0.272	−0.014	0.351	−0.067**	0.010	0.221***	0.000	−0.009	0.897	0.076***	0.000	−0.087***	0.000	0.100***	0.000	−0.026	0.413	0.040**	0.012	−0.154**	0.011
Secondary x Urban	0.085**	0.012	−0.025	0.595	−0.004	0.866	−0.050	0.175	0.294***	0.000	−0.081	0.248	0.096***	0.000	−0.074***	0.000	0.127**	0.012	−0.067	0.240	0.056***	0.006	−0.158**	0.011
Primary x Dem. Group2	0.101	0.192	−0.092	0.360	0.002	0.924	0.104***	0.008					−0.029*	0.072	−0.041**	0.043	−0.024	0.524	0.009	0.847	−0.080***	0.000	−0.061	0.186
Primary x Dem. Group3	0.034	0.719	−0.327	0.502									0.019*	0.082	−0.028**	0.049	−0.093***	0.009	0.021	0.646	−0.077***	0.002	0.019	0.759
Primary x Dem. Group4	−0.279	0.226	0.345	0.532									0.073***	0.000	−0.060***	0.000	0.096	0.200	−0.023	0.780				
Primary x Dem. Group5	0.118*	0.065	−0.131*	0.100									0.036***	0.000	−0.033**	0.020	−0.100*	0.061	−0.022	0.746				
Secondary x Dem. Group2	−0.036	0.665	0.050	0.647	−0.051	0.149	0.161***	0.008					−0.029	0.291	−0.058*	0.073	−0.020	0.774	0.013	0.866	−0.037	0.177	−0.114**	0.036
Secondary x Dem. Group3	−0.106	0.274	−0.098	0.489									0.023	0.178	−0.033*	0.097	−0.162***	0.006	0.027	0.687	−0.132*	0.095	0.109	0.317
Secondary x Dem. Group4	−0.096*	0.063	0.186	0.103									0.098***	0.000	−0.089***	0.001	0.080	0.305	0.052	0.562				
Secondary x Dem. Group5	−0.045	0.531	0.015	0.876									0.048***	0.001	−0.019	0.306	−0.041	0.545	−0.097	0.278				
Primary x Region2	0.025	0.609	−0.196***	0.009	−0.055**	0.036	0.196***	0.000	−0.043	0.548	0.051	0.535	−0.002	0.835	0.027*	0.071	0.100***	0.009	−0.081*	0.091	−0.022	0.223	−0.237***	0.001
Primary x Region3	0.044	0.626	−0.147	0.208	−0.053**	0.029	0.163***	0.001	0.073	0.261	0.052	0.495	0.068***	0.000	−0.026*	0.075	0.024	0.726	0.055	0.490	−0.052**	0.012	−0.188**	0.012
Primary x Region4	0.116***	0.002	−0.049	0.372	−0.020	0.438	0.116**	0.027	0.011	0.864	0.058	0.457	0.058***	0.000	−0.032*	0.087	0.160***	0.001	−0.055	0.311	0.070	0.184	−0.075	0.431
Primary x Region5	0.021	0.620	−0.111	0.135	−0.059**	0.025	0.172***	0.001					−0.080***	0.000	0.078***	0.000					0.073	0.181	−0.120	0.168
Primary x Region6	−0.006	0.921	−0.006	0.934	−0.044*	0.099	0.147***	0.005					−0.065***	0.000	0.017	0.241					0.000	0.993	−0.203***	0.005
Primary x Region7	0.202***	0.003	−0.108	0.268	0.000	0.988	0.059	0.307													0.028	0.224	−0.218**	0.011
Secondary x Region2	−0.003	0.965	−0.112	0.161	−0.015	0.748	−0.063	0.423	−0.100	0.402	0.086	0.485	−0.004	0.732	0.017	0.318	0.122*	0.055	−0.072	0.322	−0.022	0.346	−0.275***	0.000
Secondary x Region3	−0.052	0.556	−0.028	0.807	0.030	0.454	−0.021	0.758	−0.049	0.572	0.092	0.331	0.076***	0.000	−0.017	0.341	0.149*	0.063	−0.044	0.628	−0.055**	0.045	−0.177**	0.024
Secondary x Region4	0.123***	0.002	−0.168**	0.013	0.080**	0.048	−0.139*	0.070	−0.028	0.625	−0.026	0.737	0.088***	0.000	−0.022	0.340	0.221***	0.001	−0.030	0.711	0.031	0.731	0.069	0.579
Secondary x Region5	0.005	0.936	−0.003	0.971	0.041	0.339	−0.131*	0.081					−0.083***	0.000	0.025	0.206					0.232***	0.000	−0.225***	0.007
Secondary x Region6	−0.065	0.432	0.026	0.805	0.059	0.222	−0.060	0.473					−0.057***	0.000	−0.021	0.231					−0.033	0.216	−0.182**	0.017
Secondary x Region7	0.063	0.418	−0.047	0.673	−0.001	0.980	0.037	0.678													−0.039	0.295	−0.208**	0.028
$\beta_1 = \beta_2$	0.189		0.431		0.846		0.113		0.937		0.717		0.918		0.736		0.811		0.766		0.023		0.160	
$\pi_{female} = \rho_{female}$	0.000		0.000		0.000		0.001		0.190		0.235		0.000		0.000		0.000		0.028		0.000		0.000	
$\pi_{urban} = \rho_{urban}$	0.244		0.186		0.697		0.662		0.444		0.508		0.024		0.207		0.502		0.370		0.288		0.845	
N_{total}	80369				134610				18459				1141715				111919				327181			
Mean Outcome	0.617				0.688				0.776				0.675				0.574				0.692			

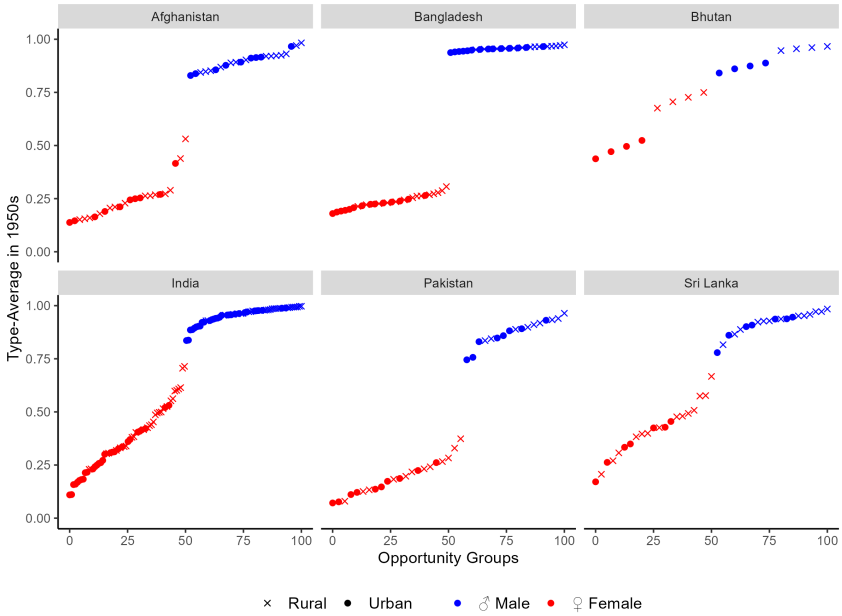
Notes: The table displays for labor force participation the coefficient estimates (Coef.) and the p-value of the t-test for their significance associated to equation (3) for the oldest and youngest cohort by country in a regression pooled across cohorts (i.e., the coefficients for the oldest cohorts are the interaction terms of the cohort dummy and the displayed baseline variable, the other cohort interaction terms are omitted from the output).
Source: Own calculations based on harmonized survey data (table A1).

Table A15: Labor Market Returns & Differentials: Wage-Employment

	Afghanistan				Bangladesh				Bhutan				India				Nepal				Pakistan				Sri Lanka			
	1960-64		1980-84		1955-59		1980-84		1960-64		1975-79		1950-54		1980-84		1950-54		1965-69		1955-59		1980-84		1950-54		1980-84	
Variable	Coef.	p-value	Coef.	p-value	Coef.	p-value	Coef.	p-value	Coef.	p-value	Coef.	p-value	Coef.	p-value	Coef.	p-value	Coef.	p-value	Coef.	p-value	Coef.	p-value	Coef.	p-value	Coef.	p-value	Coef.	p-value
(Intercept (differential Cohort FE))	0.237	0.193	−0.017	0.666	0.066	0.637	−0.062	0.382	0.554**	0.035	0.093***	0.000	0.857***	0.000	0.047***	0.004	0.590**	0.050	0.056	0.185	0.754***	0.000	0.127**	0.034	0.759***	0.000	−0.071	0.486
Age	−0.002	0.795			0.020***	0.001			−0.017	0.148			−0.012***	0.000			−0.015	0.264			−0.011	0.195			0.001	0.847		
Age^2	0.000	0.924			0.000***	0.000			0.000	0.142			0.000**	0.045			0.000	0.395			0.000	0.410			0.000	0.120		
Primary	0.162*	0.085	−0.081	0.521	−0.206***	0.000	0.162*	0.075	0.361***	0.000	−0.166	0.104	0.042*	0.065	−0.061**	0.034	0.063	0.446	−0.157	0.102	0.022	0.756	−0.124	0.147	0.005	0.836	−0.030	0.772
Secondary	0.756***	0.000	−0.122	0.520	0.251***	0.001	0.030	0.815	0.621***	0.000	−0.164	0.178	0.211***	0.000	−0.222***	0.000	0.511**	0.014	−0.234	0.359	0.236**	0.012	−0.189*	0.085	0.122***	0.000	0.017	0.875
Female	−0.076***	0.000	−0.009	0.585	−0.152***	0.000	0.031	0.292	−0.130***	0.000	−0.099***	0.000	−0.042***	0.000	−0.070***	0.000	−0.122***	0.000	−0.039	0.241	−0.058**	0.028	−0.127***	0.000	−0.015	0.254	−0.045	0.381
Urban	0.142***	0.000	−0.088***	0.007	0.148***	0.000	0.049	0.159	0.427***	0.000	−0.007	0.838	0.119***	0.000	0.052***	0.000	0.100**	0.012	0.021	0.653	0.282***	0.000	−0.135***	0.001	−0.026	0.245	−0.005	0.965
Dem. Group2	0.012	0.604	−0.016	0.649	−0.026	0.314	0.002	0.965					−0.181***	0.000	0.033*	0.054	−0.015	0.529	−0.032	0.300	−0.068	0.169	−0.002	0.973	0.324***	0.000	−0.269***	0.000
Dem. Group3	−0.022	0.431	−0.001	0.980									−0.239***	0.000	0.025*	0.081	0.083	0.119	0.083	0.300	−0.218***	0.000	0.089	0.112	0.431***	0.000	−0.256***	0.000
Dem. Group4	0.005	0.844	−0.081**	0.032									−0.260***	0.000	0.051***	0.007	0.114***	0.000	−0.024	0.515	0.154***	0.008	−0.058	0.443				
Dem. Group5	−0.012	0.419	−0.005	0.858									−0.291***	0.000	0.051***	0.002					0.070	0.167	−0.088	0.191				
Region2	−0.094***	0.000	0.014	0.631	−0.083**	0.025	0.109*	0.087	−0.052***	0.006	0.018	0.530	0.026**	0.014	−0.047***	0.005	0.005	0.864	−0.006	0.886	0.231***	0.000	−0.028	0.689	−0.126***	0.000	0.277**	0.013
Region3	−0.056*	0.067	0.042	0.340	−0.063*	0.070	0.137**	0.020	−0.011	0.658	0.040	0.258	0.177***	0.000	−0.096***	0.000	−0.002	0.941	−0.056	0.211	−0.130***	0.009	−0.067	0.330	−0.128***	0.000	0.352***	0.001
Region4	−0.092***	0.000	0.055*	0.097	−0.053	0.153	−0.007	0.907	0.008	0.681	0.013	0.652	0.128***	0.000	0.002	0.929	−0.063*	0.055	−0.011	0.810	0.091*	0.071	−0.113	0.105	−0.277***	0.000	0.364**	0.017
Region5	−0.067**	0.018	0.024	0.503	−0.021	0.578	−0.008	0.893					0.240***	0.000	−0.071***	0.000	0.010	0.817	−0.069	0.197					−0.449***	0.000	0.354**	0.012
Region6	−0.074***	0.003	0.083**	0.022	0.026	0.497	0.112*	0.087					0.262***	0.000	−0.066***	0.000									−0.305***	0.000	0.262**	0.021
Region7	−0.125***	0.000	0.015	0.700	0.051	0.258	−0.052	0.462																	−0.351***	0.000	0.133	0.321
Primary x Female	−0.202**	0.044	0.166	0.237	0.013	0.659	−0.005	0.917	−0.102	0.305	0.087	0.427	0.018	0.306	0.028	0.191	0.012	0.887	−0.065	0.494	0.267***	0.002	−0.093	0.342	−0.088***	0.000	0.059	0.268
Secondary x Female	0.329***	0.000	−0.062	0.482	0.289***	0.000	0.024	0.799	−0.047	0.631	0.284***	0.008	0.203***	0.000	0.005	0.827	−0.085	0.649	−0.010	0.963	0.257***	0.000	0.100	0.103	0.104***	0.000	−0.018	0.747
Primary x Urban	0.049	0.498	0.014	0.878	0.039	0.198	−0.088*	0.054	−0.182**	0.035	0.082	0.411	0.128***	0.000	−0.134***	0.000	0.075	0.393	−0.033	0.739	−0.193***	0.000	0.159***	0.009	0.043*	0.078	−0.017	0.871
Secondary x Urban	−0.066	0.283	−0.020	0.813	−0.066	0.134	−0.068	0.309	−0.375***	0.001	0.069	0.581	0.023	0.170	0.007	0.758	−0.125	0.401	−0.034	0.851	−0.178**	0.012	0.091	0.274	0.041	0.146	0.019	0.857
Primary x Dem. Group2	−0.052	0.613	0.158	0.256	0.065*	0.055	−0.076	0.226					0.109***	0.000	−0.096**	0.011	−0.083	0.285	0.197**	0.027	0.055	0.453	−0.022	0.807	−0.288***	0.000	0.305***	0.000
Primary x Dem. Group3	−0.091	0.509	−0.034	0.844									0.035	0.113	−0.008	0.762	0.092	0.661	−0.233	0.392	−0.001	0.986	0.013	0.886	−0.128***	0.000	0.258***	0.002
Primary x Dem. Group4	0.097	0.668	−0.250	0.288									−0.029	0.261	0.039	0.276	−0.113	0.264	0.065	0.576	−0.440***	0.000	0.306**	0.023				
Primary x Dem. Group5	−0.060	0.356	0.083	0.366									0.054***	0.003	−0.052**	0.042					−0.282***	0.001	0.232**	0.038				
Secondary x Dem. Group2	−0.131	0.337	0.040	0.839	0.050	0.359	−0.098	0.267					0.119**	0.021	−0.031	0.590	0.097	0.574	−0.005	0.980	−0.105	0.276	0.130	0.258	−0.448***	0.000	0.338***	0.000
Secondary x Dem. Group3	−0.171	0.337	0.057	0.826									0.135***	0.000	−0.076**	0.028	0.251	0.238			0.264***	0.000	−0.295***	0.002	−0.587***	0.000	0.286	0.125
Secondary x Dem. Group4	0.304*	0.063	−0.703**	0.013									0.066*	0.072	−0.021	0.637	−0.078	0.649	−0.065	0.758	−0.163	0.210	0.172	0.264				
Secondary x Dem. Group5	−0.207*	0.087	0.095	0.582									0.113***	0.000	−0.023	0.479					−0.169	0.261	0.389**	0.027				
Primary x Region2	0.067	0.614	−0.110	0.480	0.146***	0.002	−0.204**	0.014	−0.109	0.384	0.081	0.567	−0.106***	0.000	0.048*	0.080	−0.008	0.928	0.108	0.313	0.062	0.440	−0.076	0.443	0.021	0.422	−0.193*	0.090
Primary x Region3	−0.107	0.325	0.116	0.440	0.109**	0.012	−0.103	0.188	0.083	0.445	−0.038	0.761	−0.173***	0.000	0.039	0.159	0.040	0.662	0.079	0.471	0.502***	0.000	−0.251**	0.039	−0.024	0.430	−0.186*	0.099
Primary x Region4	0.084	0.325	−0.112	0.366	0.097**	0.038	−0.057	0.492	−0.084	0.456	0.012	0.925	−0.108***	0.000	−0.058	0.233	−0.036	0.683	0.189*	0.098	0.239***	0.003	−0.063	0.546	0.051	0.537	−0.242	0.129
Primary x Region5	0.008	0.935	−0.034	0.788	0.048	0.332	−0.025	0.773					−0.150***	0.000	0.058**	0.045	−0.179**	0.038	0.152	0.186					0.323***	0.001	−0.320**	0.041
Primary x Region6	−0.128*	0.077	0.015	0.886	−0.007	0.890	−0.003	0.972					−0.175***	0.000	0.090***	0.000									0.037	0.175	−0.136	0.240
Primary x Region7	−0.166	0.123	0.071	0.592	−0.033	0.557	−0.036	0.706																	0.072**	0.020	−0.031	0.823
Secondary x Region2	−0.047	0.709	0.144	0.336	−0.047	0.548	0.016	0.900	−0.243	0.157	0.311*	0.098	−0.049*	0.052	0.023	0.477	−0.044	0.795	0.190	0.377	−0.105	0.207	0.021	0.836	0.092***	0.003	−0.351***	0.002
Secondary x Region3	−0.420***	0.005	0.394*	0.059	0.050	0.472	−0.097	0.396	−0.042	0.727	0.031	0.822	−0.172***	0.000	0.106***	0.002	−0.499**	0.026	0.741***	0.007	0.233*	0.075	−0.002	0.988	0.097***	0.006	−0.377***	0.001
Secondary x Region4	−0.127	0.131	0.211*	0.081	−0.026	0.740	0.060	0.636	0.123**	0.044	−0.084	0.341	0.049	0.110	−0.103**	0.015	0.198	0.517			0.038	0.664	0.140	0.207	0.227*	0.057	−0.176	0.336
Secondary x Region5	−0.113	0.169	0.173	0.147	−0.051	0.531	−0.087	0.511					−0.203***	0.000	0.031	0.380	0.256	0.101	0.087	0.737					0.563***	0.000	−0.392**	0.039
Secondary x Region6	−0.189	0.138	−0.166	0.351	−0.156*	0.076	0.065	0.652					−0.179***	0.000	0.072**	0.021									0.228***	0.000	−0.293**	0.014
Secondary x Region7	−0.446***	0.001	0.222	0.256	−0.123	0.241	0.135	0.387																	0.273***	0.000	−0.162	0.279
$\beta_1 = \beta_2$	0.000		0.792		0.000		0.275		0.011		0.992		0.000		0.000		0.024		0.734		0.009		0.492		0.000		0.081	
$\pi_{female} = \rho_{female}$	0.000		0.167		0.000		0.731		0.654		0.145		0.000		0.340		0.629		0.806		0.931		0.115		0.000		0.004	
$\pi_{urban} = \rho_{urban}$	0.010		0.601		0.004		0.705		0.062		0.909		0.000		0.000		0.150		0.996		0.823		0.370		0.914		0.282	
N_{total}	42106				90977				14155				766510				8668				63453				222051			
Mean Outcome	0.160				0.358				0.384				0.456				0.202				0.584				0.561			

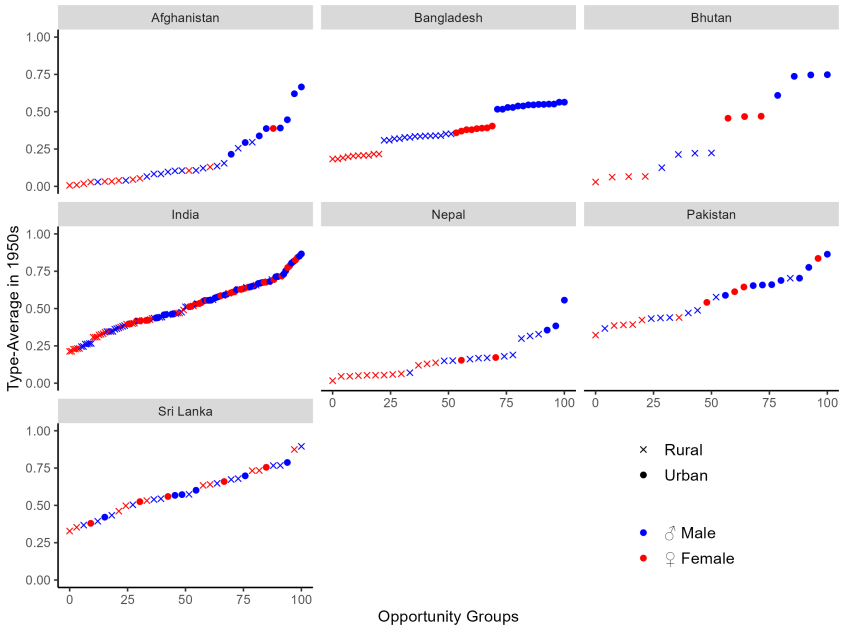
Notes: The table displays for wage employment the coefficient estimates (Coef.) and the p-value of the t-test for their significance associated to equation (4) for the oldest and youngest cohort by country in a regression pooled across cohorts (i.e., the coefficients for the oldest cohorts are the interaction terms of the cohort dummy and the displayed baseline variable, the other cohort interaction terms are omitted from the output).
Source: Own calculations based on harmonized survey data (table A1).

Figure A20: Profiles 1950s - LFP



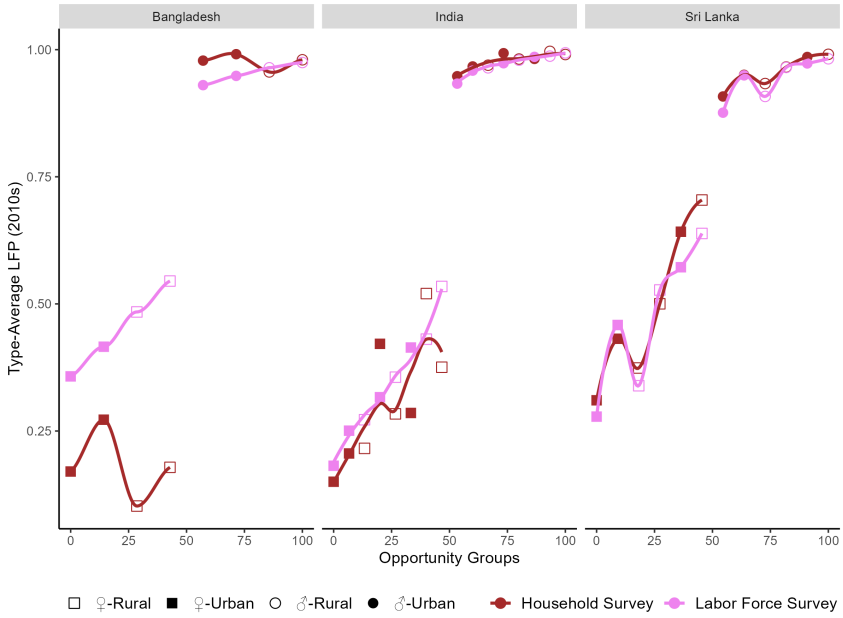
Notes: The figure depicts the opportunity profiles for labor force participation of the 1950s birth cohort across countries (see figure A4 for detailed explanation). Group characteristics (circumstances) are indicated by the symbol (Urbanity) and its color (Gender), e.g., red crosses indicate females living in rural areas. We note the staggering differences between groups, especially men and women.
Source: Own calculations based on harmonized survey data (table A1).

Figure A21: Profiles 1950s - Wage-Employment



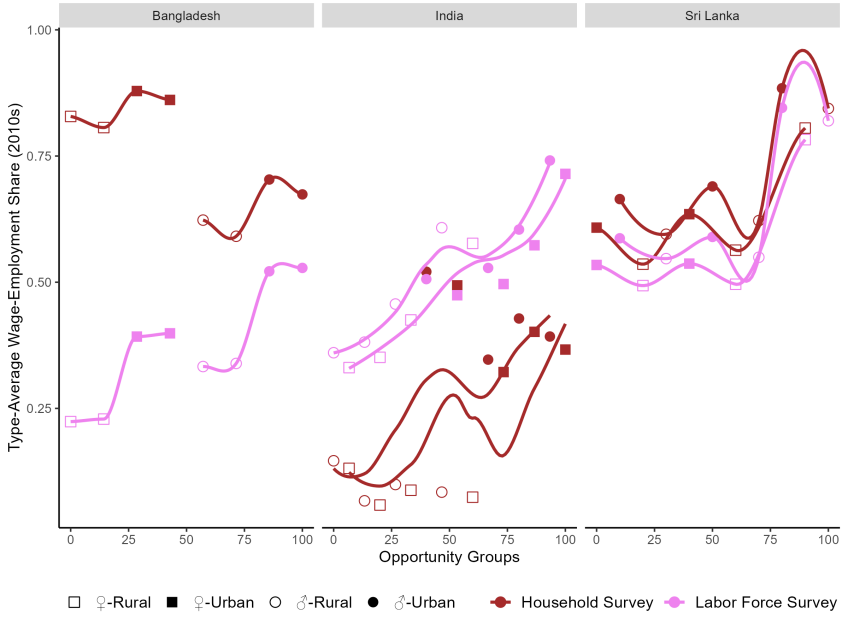
Notes: The figure depicts the opportunity profiles for wage employment of the 1950s birth cohort across countries (see figure A4 for detailed explanation). Group characteristics (circumstances) are indicated by the symbol (Urbanity) and its color (Gender), e.g., red crosses indicate females living in rural areas. We note the staggering differences between groups, especially men and women.
Source: Own calculations based on harmonized survey data (table A1).

Figure A22: Profiles LFP across Surveys (2010s)



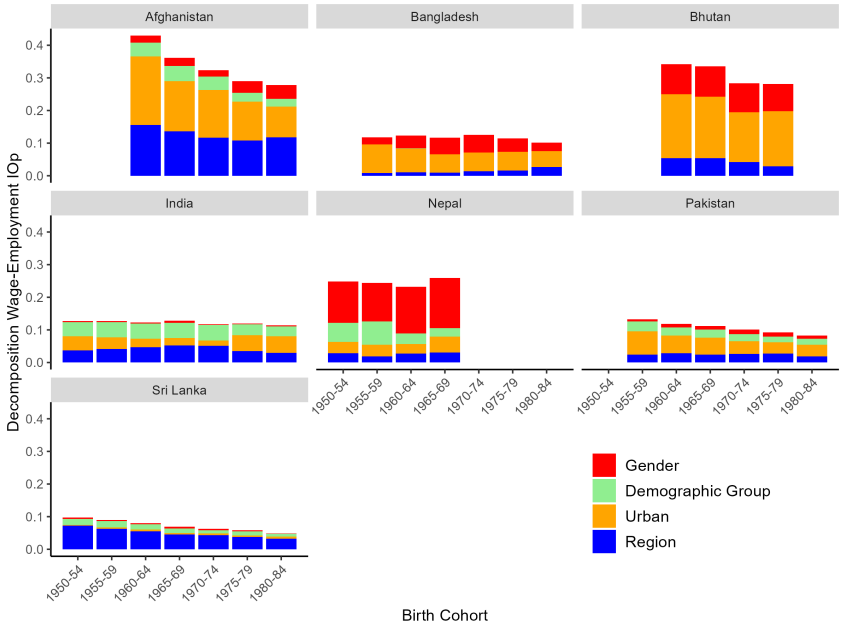
Notes: The figure depicts, in the latest cross-section where the both potential data sources, household surveys (brown) and labor force surveys (pink), are available (India 2011, Bangladesh/Sri Lanka 2015/2016), the opportunity profiles for labor force participation, i.e., the predicted average LFP by type (opportunity group) where the type definition only relies on urbanity (urban vs. rural; full vs. empty symbol), gender (female vs. male; squares vs. circles) and demographic group. Opportunity groups ordering (x-axis) is based to their predicted LFP (y-axis) in the labor force survey (our preferred data source). We note (i) a considerable divergence in female LFP for Bangladesh across sources, i.e., the labor force survey oversamples active women compared to the household survey, (ii) for India and Sri Lanka, estimates of both data sources align closely.
Source: Own calculations based on harmonized survey data (table A1).

Figure A23: Profiles Wage-Employment across Surveys (2010s)



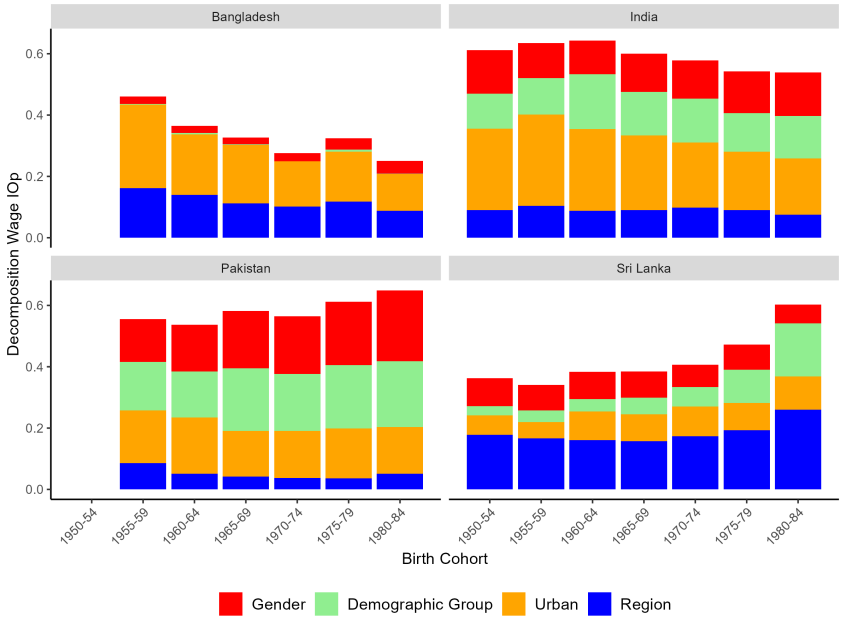
Notes: The figure depicts, analogous to figure A22, the opportunity profiles for *wage employment*.
We note (i) for Bangladesh, in the household survey, a non-reasonable share of female wage-employment of around 75 percent which discredits its usage and, hence, triggers the exclusion of wage-employment estimates based on the household survey from the robustness checks, (ii) for India, the systematically higher prevalence of wage-employment in the labor force survey, and (iii) for Sri Lanka, estimates of both data sources align more closely.
Source: Own calculations based on harmonized survey data (table A1).

Figure A24: Decomposing Wage-Employment IOp by Circumstances



Notes: The figure depicts the importance of the different circumstances for wage employment IOp as obtained via a Shapley decomposition (see footnote 30). While the hight of the columns represents the value the D-Index, its colored segments indicate the contribution of the corresponding circumstance to it. No systematic substantial changes in the composition of IOp can be observed.
Source: Own calculations based on harmonized survey data (table A1).

Figure A25: Decomposing Wage IOp by Circumstances



Notes: The figure depicts the importance of the different circumstances for wage IOp as obtained via a Shapley decomposition (see footnote 30). While the hight of the columns represents the extend of relative IOp, its colored segments indicate the contribution of the corresponding circumstance to it. No systematic substantial changes in the composition of IOp can be observed.
Source: Own calculations based on harmonized survey data (table A1).