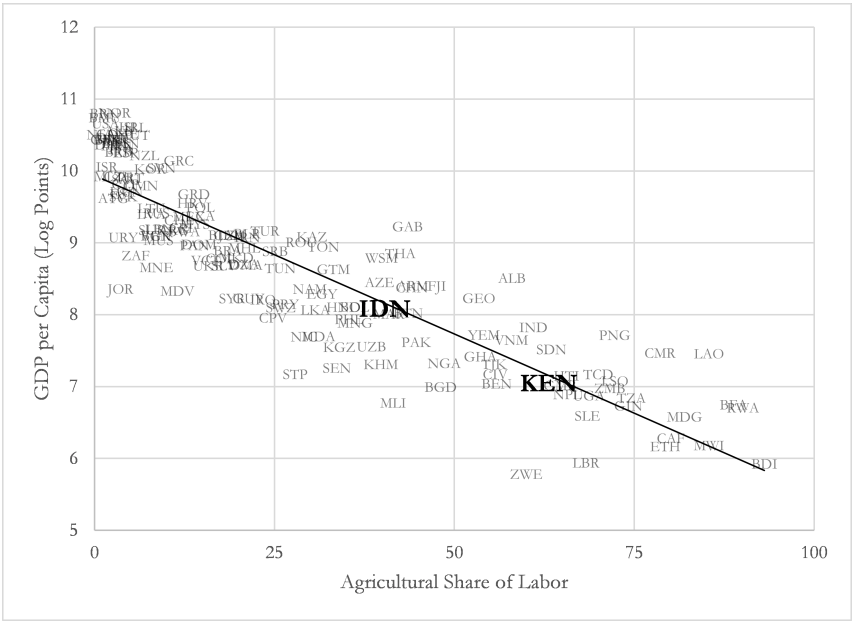
Supplementary Appendix Materials (for online publication)

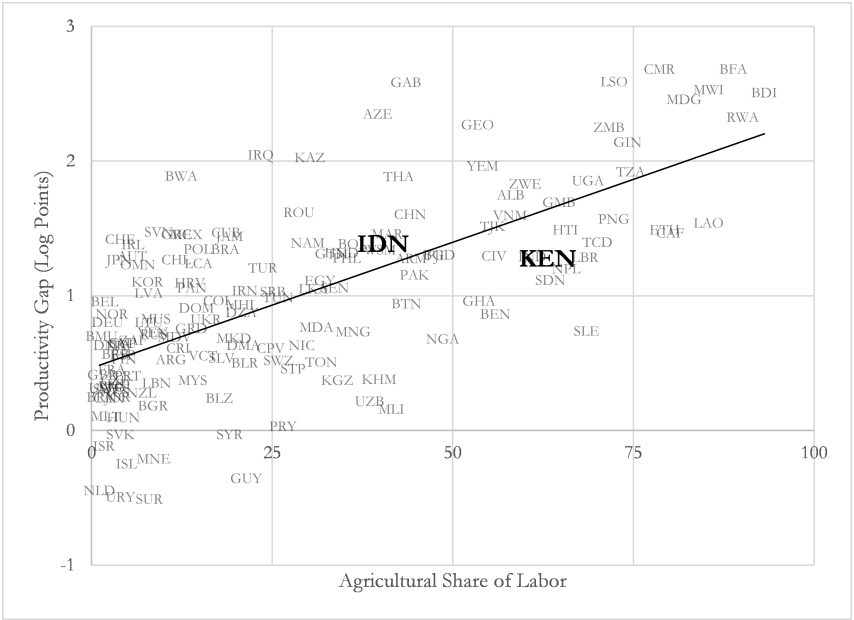
Appendix A: Additional Figures and Tables

Figure A1: Log GDP per Capita and Agricultural Share



Notes: Table source data is from Gollin, Lagakos, and Waugh (2014), Online Appendix Table A4. Kenya (KEN) and Indonesia (IDN) are highlighted.

Figure A2: Agricultural Share and Agricultural Productivity Gap



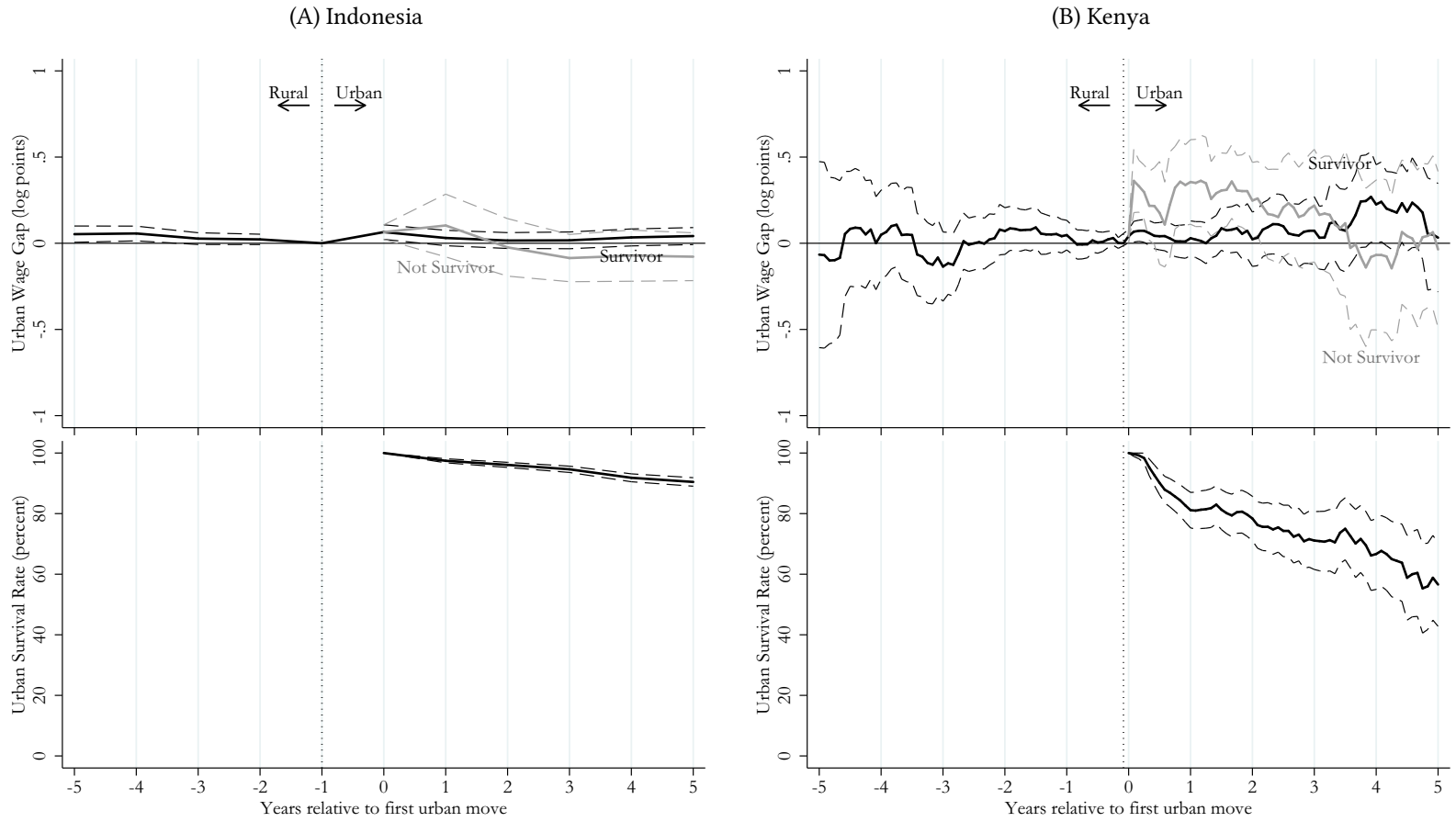
Notes: Table source data is from Gollin, Lagakos, and Waugh (2014), Online Appendix Table A4. Kenya (KEN) and Indonesia (IDN) are highlighted.

Figure A3: Types of Individual Agricultural Productivity Data

	Lower quality measures ←			→ Higher quality measures
(A) Indonesia				
Source of agricultural productivity and hours		**Self-employed profits (commercial and subsistence agriculture)[1]**		**Wage employment**
Individual-years in Agriculture		55,130		29,155
Individuals in Agriculture		6,867		5,666
Agriculture productivity gap		0.128***		–0.019
(Standard error)		(0.030)		(0.024)
[Individual-years]		[134,153]		[139,846]
		0.077***		
		(0.020)		
		258,745		
(B) Kenya				
Source of agricultural productivity and hours	**Less reliable individual agricultural productivity data[2]**	**Self-employed profits (subsistence agriculture)**	**Self-employed profits (commencial agriculture)**	**Wage employment**
Individual-months in Agriculture	3,507	2,331	4,225	13,754
Individuals in Agriculture	348	205	137	537
Agriculture productivity gap		0.031		0.098
(Standard error)		(0.177)		(0.120)
[Individual-months]		[37,064]		[94,653]
		0.014		
		(0.106)		
		130,322		

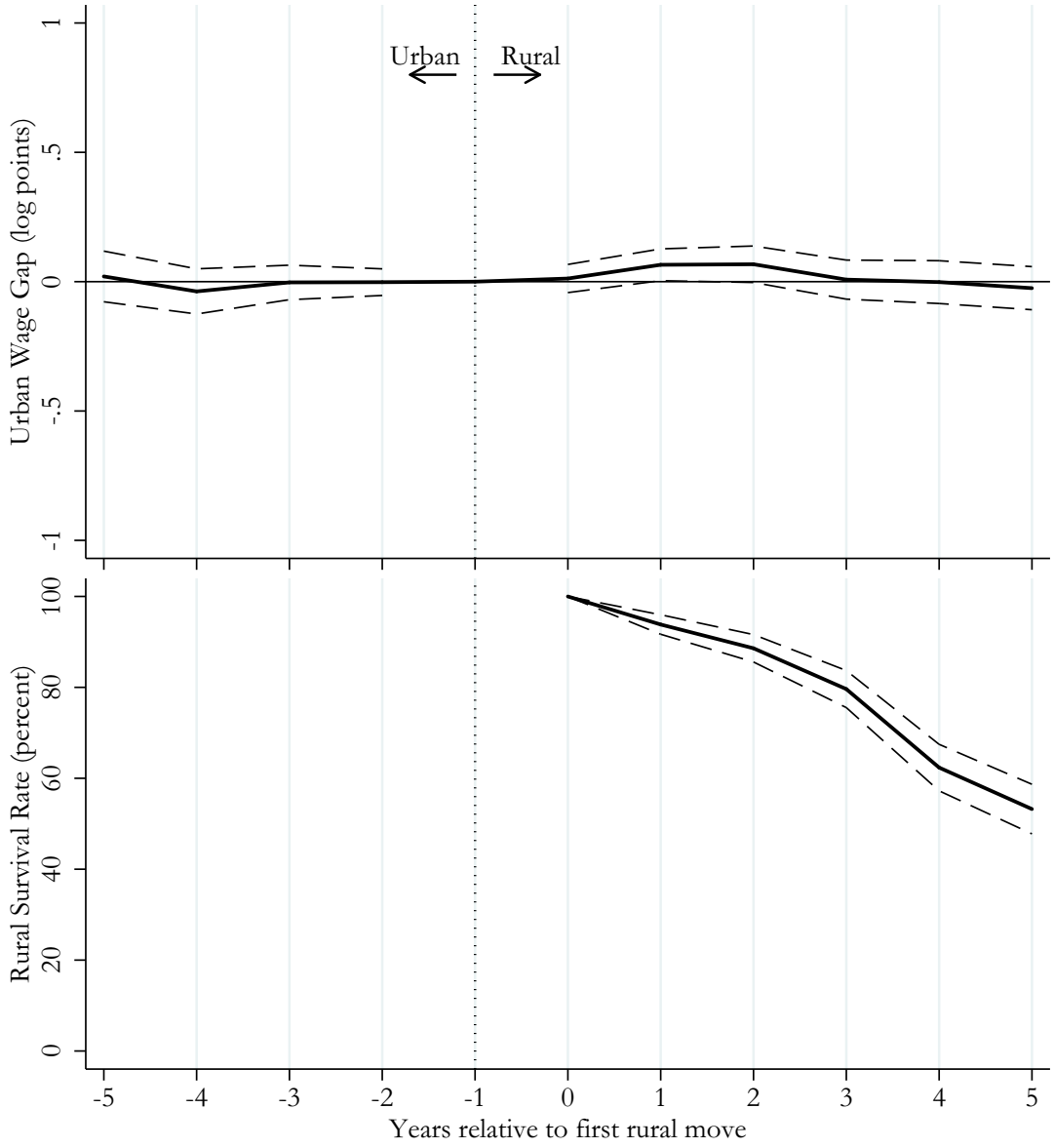
Notes: [1]The IFLS does not distinguish between profits in subsistence and commercial agriculture. [2]Less reliable agricultural productivity data encompasses individual-months where the only source of agricultural productivity data is from activities where the respondent is not the main decision maker and other household members contribute some hours. All estimates in this figure are based on log wages as the outcome variable and can be found in Appendix Table A10.

Figure A4: Event Study of Urban Migration for Urban Survivors



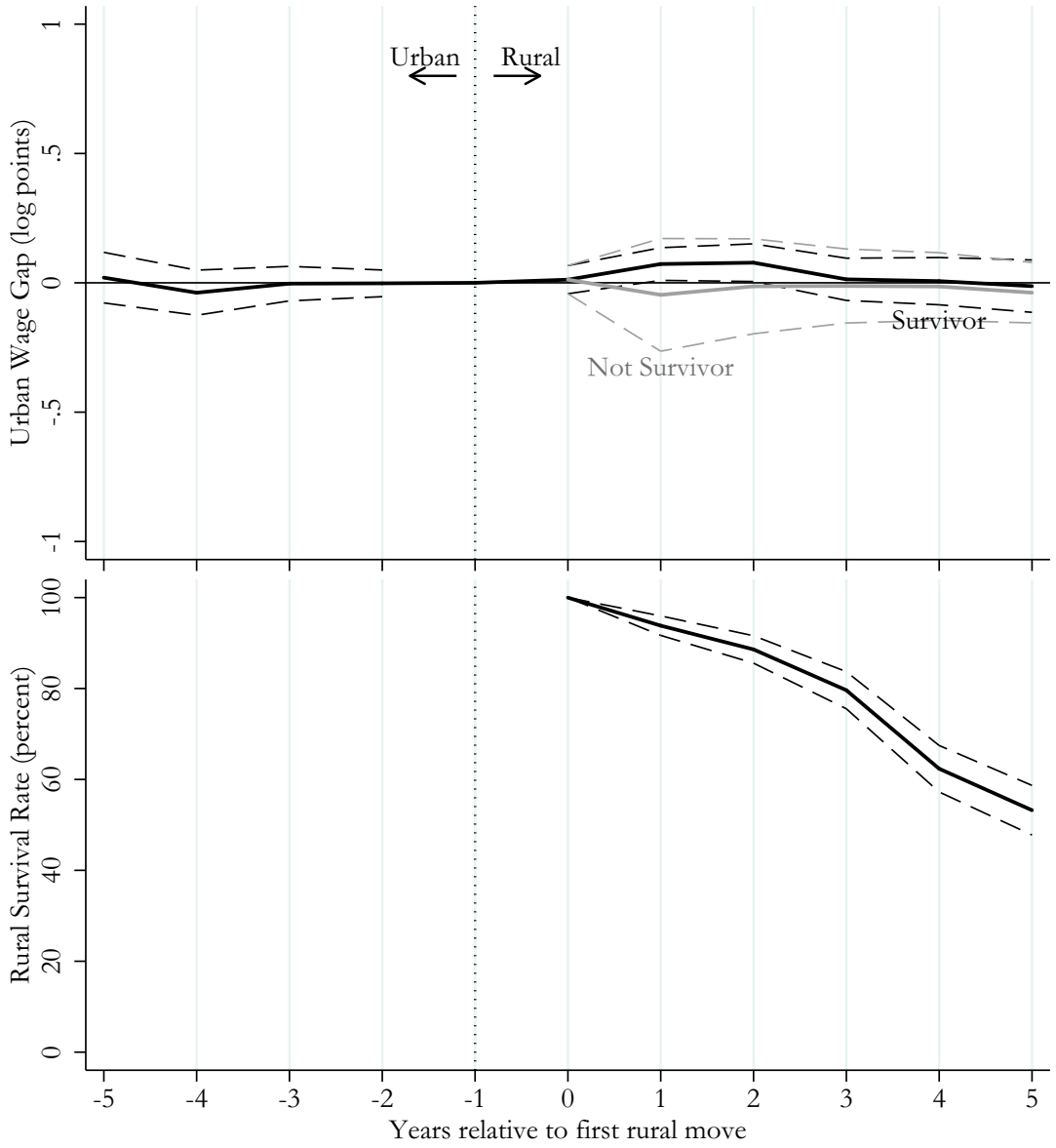
Notes: Event study coefficients reported in top half of figure separately for "survivors" and "not-survivors." "Survivor" status is defined as having no rural observations from period zero (when the individual moved an urban area) to the period of interest, corresponding exactly to the survivor rate graph on the lower half of the figure. Survivor coefficients (black line in the top half) obtained by interacting a survivor indicator with post-event time indicators described in Section 4B; "not-survivor" coefficients (grey line in the top half) is the event time indicator interacted with one minus the survivor indicator. Panel A reports results for Indonesia, and Panel B reports results for Kenya. Please refer to Figure 3 notes for additional details on included control variables and computation of survivor rates.

Figure A5: Event Study of Rural Migration



Notes: Figure uses data on individuals in the IFLS who are born in urban areas. Event time indicator variables defined analogously to Figure 3 except with respect to individuals' first observed rural move. Coefficients multiplied by negative 1 to interpret difference in earnings as an urban premium. Sample includes 1,296 movers with wage observations at the time of move and one period prior; 636 individuals report wages five years later. Please refer to Figure 3 notes for additional details on included control variables and computation of survivor rates.

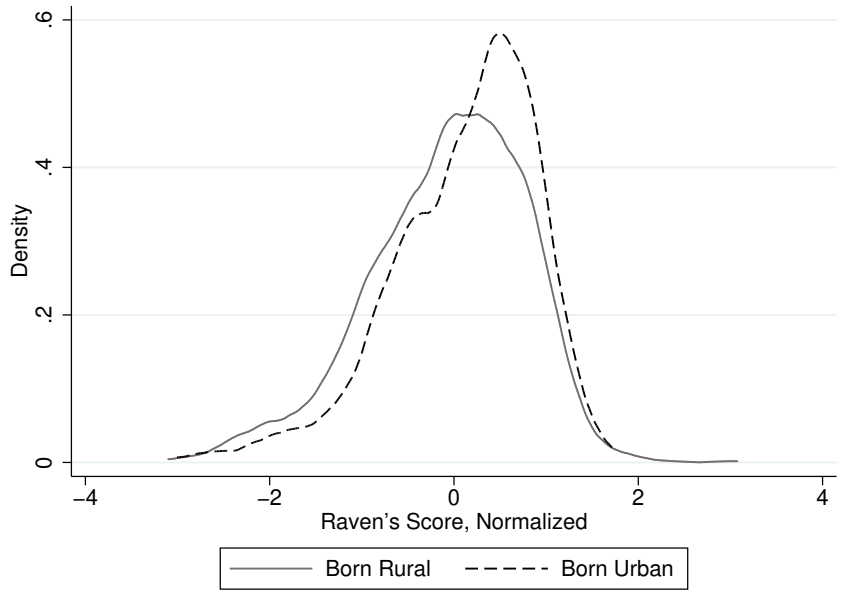
Figure A6: Event Study of Rural Migration for Survivors



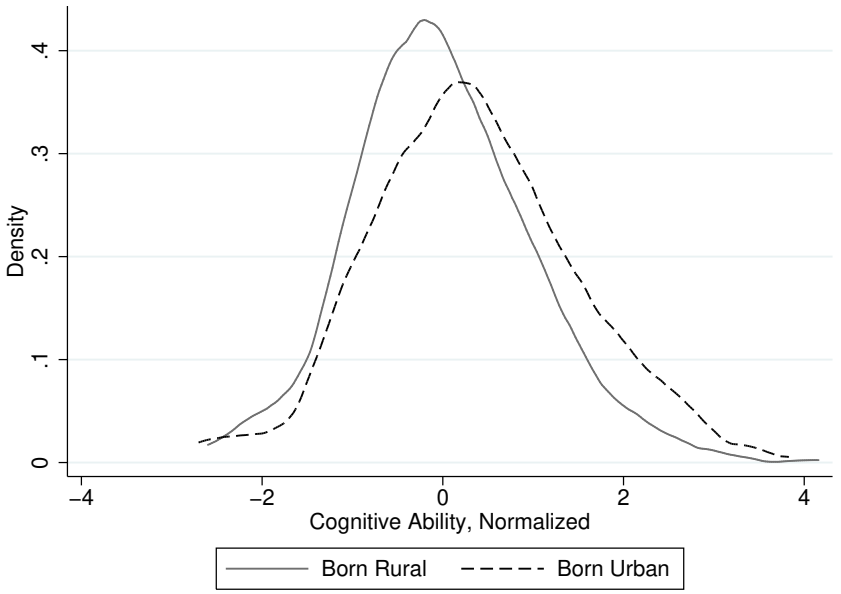
Notes: Figure uses data on individuals in the IFLS who are born in urban. Event study coefficients reported in top half of figure separately for "survivors" and "not-survivors." "Survivor" status is defined as having no urban observations from period zero (when the individual moved a rural area) to the period of interest, corresponding exactly to the survivor rate graph on the lower half of the figure. Survivor coefficients (black line in the top half) obtained by interacting a survivor indicator with post-event time indicators described in Section 4B; "not-survivor" coefficients (grey line in the top half) is the event time indicator interacted with one minus the survivor indicator. Panel A reports results for Indonesia, and Panel B reports results for Kenya. Please refer to Figure 3 notes for additional details on included control variables and computation of survivor rates.

Figure A7: Marginal Distributions of Cognitive Ability

(A) Indonesia—Normalized Ravens Matrices



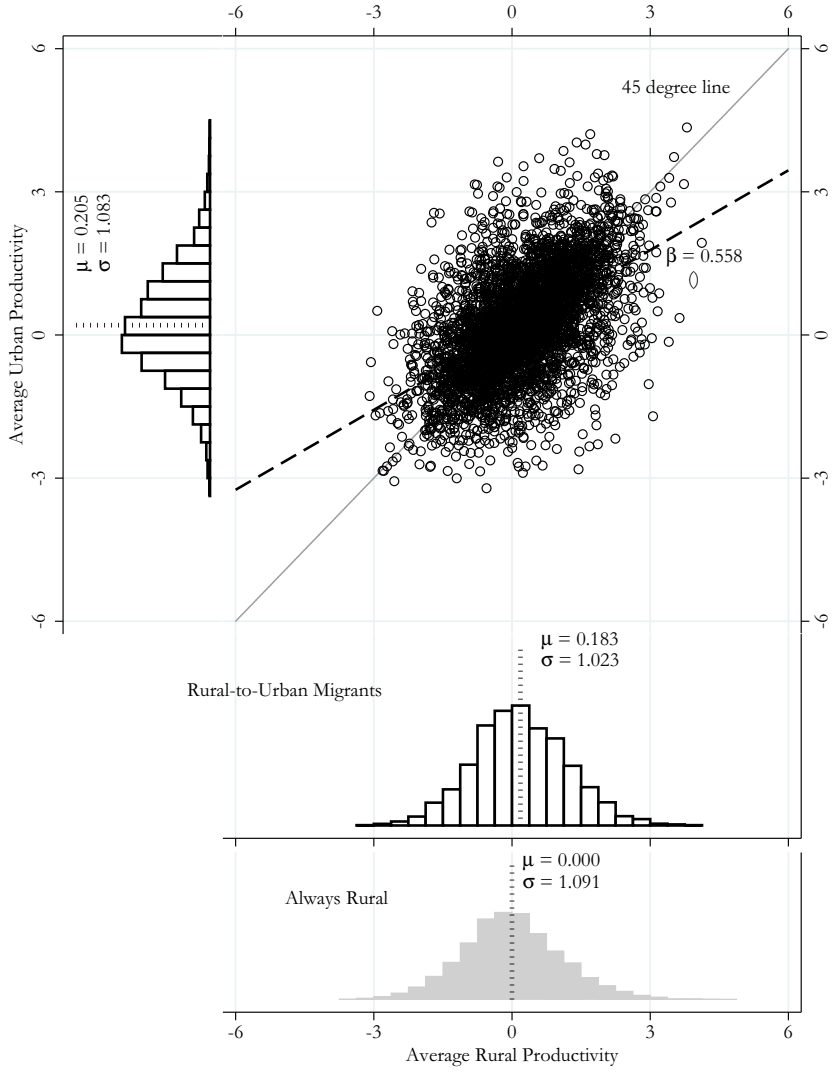
(B) Kenya—Normalized Cognitive Ability Index



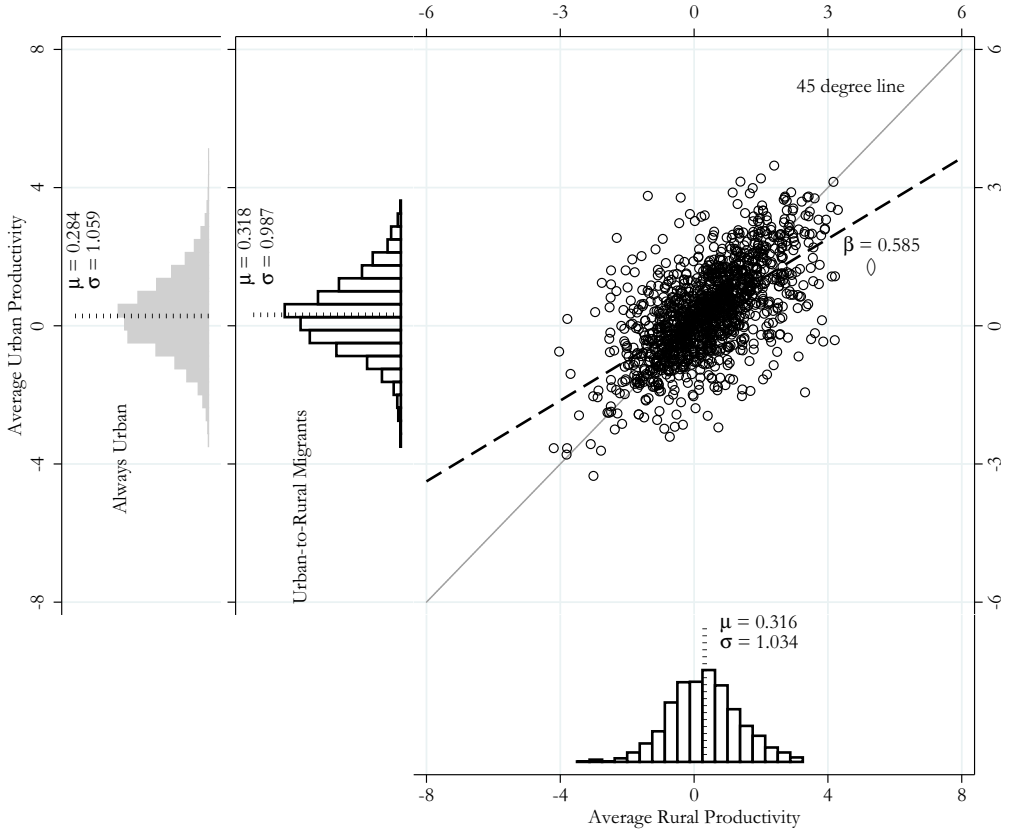
Notes: Panel A uses data from the IFLS, and Panel B uses data from the KLPS Kids sample of children of the KLPS sample, age 4–6. Panel A shows the marginal distributions of Raven matrix scores, normalized by one-year age bins, and Panel B shows the marginal distribution of a constructed Cognitive Ability Index, normalized by six-month age bins. See Table 7 for additional details about the cognitive ability index.

Figure A8: Joint Distribution of Rural and Urban Productivities

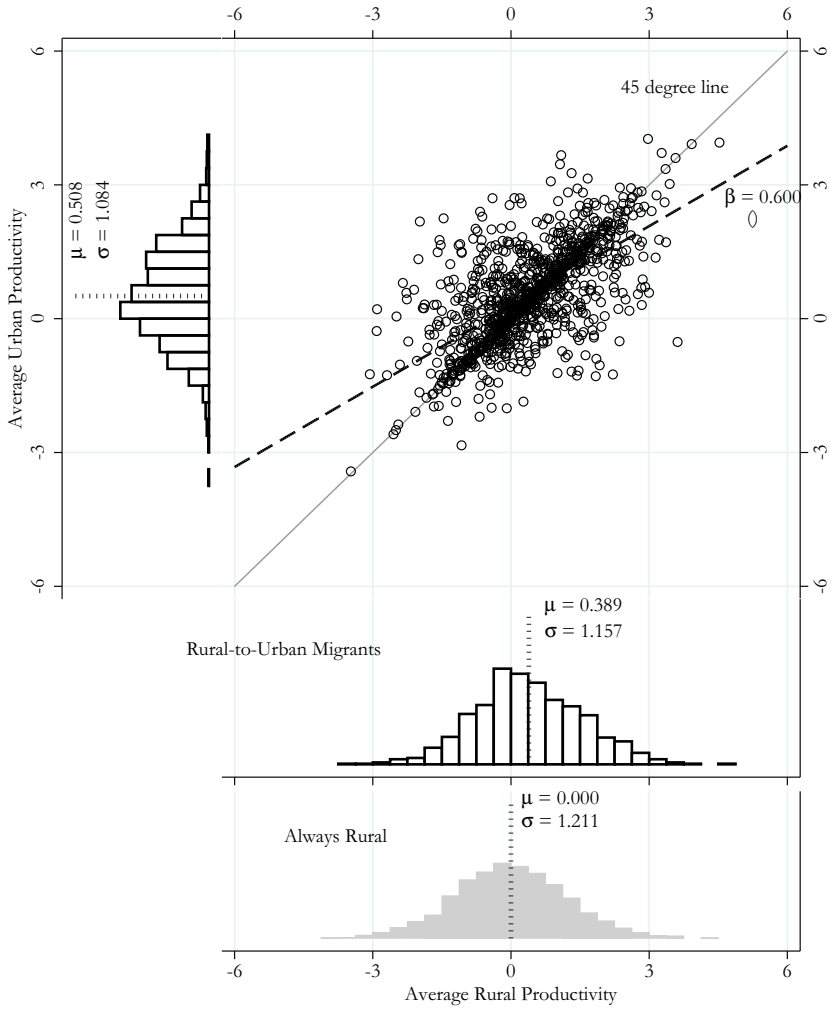
(A) Indonesia (Born Rural)



(B) Indonesia (Born Urban)



(C) Kenya (Born Rural)



Notes: Productivities are recovered individual-urban status effects from a fixed effects regression of log wages on squared age and indicators for time period on the same sample used in Tables 4 and 5. Productivities are normalized such that the average productivity of rural non-migrants has zero mean. Histograms on the bottom of Panel A represent marginal distributions of rural productivities for "Always Rural" non-migrants (grey) and migrants (hollow). Marginal distribution of estimated urban productivities for migrants reported on the left (hollow). Means and standard deviations reported in log points. Scatterplot presents joint distribution for migrants with best fit line. Bootstrapped standard error of the slope reported in parentheses from 1,000 iterations of block sampling of individuals with replacement. Panel B presents a histogram of "Always Urban" urban productivities of non-migrants (grey) at the top left, an adjacent histogram of migrant urban productivities (hollow), and migrant rural productivities (grey) below. Joint distribution of urban and rural productivities and corresponding best fit line presented similar to panel A. Panel C mimics the format of Panel A except uses data from the KLPS.

Table A1: Correlates of Employment in Non-Agriculture

(A) Indonesia (Born Rural)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Primary Ed.	0.285***					0.212***	0.236***
	(0.009)					(0.013)	(0.009)
Secondary Ed.		0.226***				0.131***	0.162***
		(0.005)				(0.007)	(0.006)
College			0.225***			0.0513***	0.0649***
			(0.005)			(0.007)	(0.006)
Female				0.0562***		0.0821***	0.0884***
				(0.006)		(0.006)	(0.005)
Raven's Z-score					0.0643***	0.0358***	
					(0.004)	(0.004)	
Constant	0.518***	0.687***	0.741***	0.734***	0.792***	0.514***	0.466***
	(0.009)	(0.004)	(0.003)	(0.004)	(0.003)	(0.013)	(0.009)
Observations	21808	21808	21808	21808	16041	16041	21808

(B) Kenya (Born Rural)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Primary Ed.	0.140***					0.103***	0.114***
	(0.012)					(0.014)	(0.013)
Secondary Ed.		0.103***				0.0451***	0.0552***
		(0.007)				(0.008)	(0.008)
College			0.0909***			0.0152	0.0245**
			(0.008)			(0.008)	(0.008)
Female				0.0194*		0.0314***	0.0279***
				(0.009)		(0.009)	(0.008)
Raven's Z-score					0.0452***	0.0214***	
					(0.004)	(0.005)	
Constant	0.803***	0.870***	0.903***	0.896***	0.905***	0.796***	0.787***
	(0.011)	(0.006)	(0.004)	(0.006)	(0.004)	(0.013)	(0.012)
Observations	4718	4718	4718	4718	4452	4452	4718

Notes: See Table 2 for sample restrictions and row variable definitions. Each cell reports a regression coefficient with an indicator for being ever being employed in non-agriculture as the dependent variable. Panel A (Indonesia) is estimated on individuals who are born in rural areas, whereas panel B (Kenya) includes the full sample subject to previously defined sample restrictions. Columns 6 and 7 report coefficients from a multiple regression with corresponding rows as included covariates. Column 7 omits the Raven's matrix exam to preserve sample size. Robust standard errors reported below in parentheses, *** p<0.01, ** p<0.05, * p<0.1.

Table A2: Correlates of Urban Migration

(A) Indonesia (Born Rural)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Primary Ed.	0.199***					0.124***	0.155***
	(0.009)					(0.013)	(0.009)
Secondary Ed.		0.168***				0.0898***	0.119***
		(0.007)				(0.009)	(0.008)
College			0.180***			0.0391*	0.0652***
			(0.013)			(0.016)	(0.014)
Female				0.0139*		0.0365***	0.0358***
				(0.007)		(0.008)	(0.007)
Raven's Z-score					0.0668***	0.0471***	
					(0.004)	(0.004)	
Constant	0.286***	0.401***	0.440***	0.447***	0.468***	0.305***	0.265***
	(0.008)	(0.004)	(0.003)	(0.004)	(0.004)	(0.012)	(0.009)
Observations	21808	21808	21808	21808	16041	16041	21808

(B) Kenya (Born Rural)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Primary Ed.	0.167***					0.0859***	0.108***
	(0.016)					(0.019)	(0.018)
Secondary Ed.		0.168***				0.0988***	0.112***
		(0.014)				(0.017)	(0.016)
College			0.231***			0.114***	0.127***
			(0.026)			(0.028)	(0.028)
Female				0.00295		0.0180	0.0147
				(0.014)		(0.014)	(0.014)
Raven's Z-score					0.0669***	0.0310***	
					(0.007)	(0.008)	
Constant	0.538***	0.601***	0.652***	0.659***	0.659***	0.548***	0.530***
	(0.014)	(0.009)	(0.007)	(0.010)	(0.007)	(0.017)	(0.016)
Observations	4718	4718	4718	4718	4452	4452	4718

Notes: See Table 2 for sample restrictions and row variable definitions. Each cell reports a regression coefficient with an indicator for being an urban migrant as the dependent variable. Both panels are estimated on individuals who are born rural. Columns 6 and 7 report coefficients from multiple regressions with corresponding rows as included covariates. Column 7 omits the Raven's matrix exam to preserve sample size. Robust standard errors reported below in parentheses, *** p<0.01, ** p<0.05, * p<0.1.

Table A3: Correlates of Employment in Non-Agriculture—Indonesia (Born Urban)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Primary Ed.	0.198***					0.114***	0.150***
	(0.020)					(0.027)	(0.020)
Secondary Ed.		0.101***				0.0774***	0.0805***
		(0.006)				(0.007)	(0.006)
College			0.0602***			0.00357	0.00913*
			(0.004)			(0.004)	(0.004)
Female				0.0293***		0.0370***	0.0333***
				(0.005)		(0.005)	(0.005)
Raven's Z-score					0.0306***	0.0153***	
					(0.004)	(0.004)	
Constant	0.749***	0.878***	0.926***	0.925***	0.937***	0.764***	0.731***
	(0.020)	(0.005)	(0.003)	(0.004)	(0.003)	(0.027)	(0.020)
Observations	9718	9718	9718	9718	7163	7163	9718

Notes: This table is analogous to Appendix Table A1 but is estimated on individuals born in urban areas in Indonesia. Please see the notes from Appendix Table A1.

Table A4: Correlates of Rural Migration—Indonesia (Born Urban)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Primary Ed.	-0.240***					-0.177***	-0.181***
	(0.023)					(0.033)	(0.024)
Secondary Ed.		-0.126***				-0.101***	-0.104***
		(0.009)				(0.012)	(0.010)
College			-0.0673***			0.0155	-0.00186
			(0.011)			(0.014)	(0.011)
Female				-0.0465***		-0.0542***	-0.0516***
				(0.009)		(0.010)	(0.009)
Raven's Z-score					-0.0213***	-0.00271	
					(0.006)	(0.006)	
Constant	0.484***	0.330***	0.269***	0.277***	0.266***	0.519***	0.513***
	(0.023)	(0.007)	(0.005)	(0.006)	(0.005)	(0.032)	(0.023)
Observations	9718	9718	9718	9718	7163	7163	9718

Notes: This table is a rural migration analog of Appendix Table A2 for Indonesia. Each cell represents a regression coefficient with an indicator for being a rural migrant as the dependent variable. The sample is restricted to individuals born in urban areas. Please see the notes from Appendix Table A2.

Table A5: Kenya Urban Towns

	Population	Percentage of Urban Individual-Months
Nairobi	3,133,518	43.2
Mombasa	938,131	14.6
Busia	61,715	6.8
Nakuru	307,990	4.5
Kisumu	409,928	4.4
Eldoret	289,380	2.5
Kakamega	91,768	1.3
Bungoma	81,151	1.1
Kitale	106,187	1.1
Naivasha	181,966	0.9
Gilgil	35,293	0.5
Other	.	19.0

Notes: This table presents a list of reported towns from urban individual-month observations. Urban status is defined based on respondent answering that they live in a large town or city. The final column lists the fraction of individual months in analysis from a particular town. The source for town populations is the 2009 Kenya Census.

Table A6: Non-Agricultural/Agricultural Gap in Earnings using Alternative Definition of Agriculture

(A) Indonesia

	Dependent variable: Log Earnings							
	(1)	(2)	(3)	(4) Log Wage	(5)	(6)	(7)	(8) Log Wage
Only non-agricultural employment	0.636***	0.331***	0.104***	0.040**				
	(0.012)	(0.011)	(0.015)	(0.018)				
Any non-agricultural employment					0.734***	0.379***	0.265***	0.098***
					(0.013)	(0.012)	(0.016)	(0.019)
Individual fixed effects	N	N	Y	Y	N	N	Y	Y
Control variables and time FE	N	Y	Y	Y	N	Y	Y	Y
Number of observations	258745	258745	258745	258745	258745	258745	258745	258745
Number of individuals	31537	31537	31537	31537	31537	31537	31537	31537

(B) Kenya

	Dependent variable: Log Earnings							
	(1)	(2)	(3)	(4) Log Wage	(5)	(6)	(7)	(8) Log Wage
Only non-agricultural employment	0.688***	0.452***	0.155*	0.057				
	(0.058)	(0.052)	(0.079)	(0.096)				
Any non-agricultural employment					0.752***	0.495***	0.287***	0.010
					(0.061)	(0.055)	(0.087)	(0.108)
Individual fixed effects	N	N	Y	Y	N	N	Y	Y
Control variables and time FE	N	Y	Y	Y	N	Y	Y	Y
Number of observations	130322	130322	130322	130322	130322	130322	130322	130322
Number of individuals	4718	4718	4718	4718	4718	4718	4718	4718

Notes: Panel A uses data from the IFLS, and Panel B uses data from the KLPS. The table repeats some of the analyses shown in Tables 4 and 5 with alternate definitions of non-agriculture. In the first "Only non-agricultural employment," an individual-time is considered *agricultural* if any of their jobs are *agricultural*, and non-agricultural otherwise. In the second, "Any non-agricultural employment," an individual-time is considered *non-agricultural* if any of their jobs are *non-agricultural*, and agricultural otherwise. For columns 4 and 8, the dependent variable is the log of earnings divided by hours worked. Control variables include log hours, log hours squared, age, age squared, years of education, years of education squared and an indicator for being female. When also including individual fixed effects in columns 3, 4, 7 and 8, the control variables are reduced to only age squared. All regressions are clustered at the individual level. Robust standard errors are in parentheses, *** p<0.01, ** p<0.05, * p<0.1.

Table A7: Non-Agricultural/Agricultural Gap in Earnings Within Rural Areas

(A) Indonesia

	Dependent variable: Log Earnings			
	(1)	(2)	(3)	(4) Log Wage
Non-agricultural employment	0.550***	0.326***	0.228***	0.074***
	(0.015)	(0.014)	(0.021)	(0.025)
Individual fixed effects	N	N	Y	Y
Control variables and time FE	N	Y	Y	Y
Number of observations	133726	133726	133726	133726
Number of individuals	18778	18778	18778	18778

(B) Kenya

	Dependent variable: Log Earnings			
	(1)	(2)	(3)	(4) Log Wage
Non-agricultural employment	0.326***	0.171***	−0.057	−0.299**
	(0.068)	(0.061)	(0.115)	(0.143)
Individual fixed effects	N	N	Y	Y
Control variables and time FE	N	Y	Y	Y
Number of observations	61750	61750	61750	61750
Number of individuals	2889	2889	2889	2889

Notes: Panel A uses data from the IFLS, and Panel B uses data from the KLPS. The table repeats some of the analyses shown in Table 5, but restricts the sample to observations where the individual resides in rural areas. For column 4, the dependent variable is the log of earnings divided by hours worked. Control variables include log hours, log hours squared, age, age squared, years of education, years of education squared and an indicator for being female. When also including individual fixed effects in columns 3 and 4, the control variables are reduced to only age squared. All regressions are clustered at the individual level. Robust standard errors are in parentheses, *** p<0.01, ** p<0.05, * p<0.1.

Table A8: Non-Agricultural/Agricultural Gap in Hours Worked

(A) Indonesia

	Dependent variable: Log Hours				
	(1)	(2)	(3)	(4)	(5)
Non-agricultural employment	0.263***	0.275***	0.265***	0.210***	0.195***
	(0.009)	(0.009)	(0.011)	(0.016)	(0.017)
Female		−0.239***	−0.231***	−0.270***	
		(0.009)	(0.010)	(0.024)	
Years of education		0.020***	0.020***	0.014*	
		(0.003)	(0.004)	(0.007)	
Years of education squared		−0.001***	−0.001***	−0.000	
		(0.000)	(0.000)	(0.000)	
Normalized Ravens			0.010**		
			(0.005)		
Normalized Ravens squared			0.002		
			(0.004)		
Individual fixed effects	N	N	N	N	Y
Time fixed effects	N	Y	Y	Y	Y
Switchers only				Y	
Number of observations	258745	258745	196354	48479	258745
Number of individuals	31537	31537	23214	3907	31537

(B) Kenya

	Dependent variable: Log Hours				
	(1)	(2)	(3)	(4)	(5)
Non-agricultural employment	0.395***	0.417***	0.432***	0.343***	0.319***
	(0.043)	(0.044)	(0.045)	(0.083)	(0.077)
Female		−0.169***	−0.165***	−0.625***	
		(0.027)	(0.028)	(0.126)	
Years of education		−0.017	−0.018	−0.080	
		(0.029)	(0.030)	(0.107)	
Years of education squared		0.002	0.002	0.004	
		(0.001)	(0.001)	(0.006)	
Normalized Ravens			0.033**	0.070	
			(0.016)	(0.062)	
Normalized Ravens squared			−0.019	−0.078*	
			(0.013)	(0.042)	
Individual fixed effects	N	N	N	N	Y
Time fixed effects	N	Y	Y	Y	Y
Switchers only				Y	
Number of observations	130322	130322	124481	14345	130322
Number of individuals	4718	4718	4452	324	4718

Notes: Panel A uses data from the IFLS and Panel B uses data from the KLPS. Please refer to Section 3 for further details on the data and to the notes of Tables 4 and 5 for additional information on the variables. The dependent variable is log hours worked in wage and self-employment. If an individual has multiple jobs in the same time period, hours from all employment are included. The sample size in column 3 is smaller in Panel A because the Raven's test was administered only for a subset of the Indonesian sample. The sample size in column 4 is smaller because it only includes "switchers" who have at least one observation in both the non-agricultural and agricultural sector. Each regression in columns 2–5 include quadratic controls for age. All regressions are clustered at the individual level. Robust standard errors are in parentheses, *** p<0.01, ** p<0.05, * p<0.1.

Table A9: Urban/Rural Gap in Hours Worked

(A) Indonesia

	Dependent variable: Log Hours				
	(1)	(2)	(3)	(4)	(5)
Urban	0.133***	0.123***	0.102***	0.055***	0.006
	(0.007)	(0.008)	(0.009)	(0.012)	(0.012)
Female		−0.208***	−0.199***	−0.193***	
		(0.009)	(0.010)	(0.018)	
Years of education		0.026***	0.026***	0.030***	
		(0.003)	(0.004)	(0.007)	
Years of education squared		−0.001***	−0.001***	−0.001***	
		(0.000)	(0.000)	(0.000)	
Normalized Ravens			0.017***		
			(0.005)		
Normalized Ravens squared			0.001		
			(0.004)		
Individual fixed effects	N	N	N	N	Y
Time fixed effects	N	Y	Y	Y	Y
Switchers only				Y	
Number of observations	258745	258745	196354	69519	258745
Number of individuals	31537	31537	23214	5683	31537

(B) Kenya

	Dependent variable: Log Hours				
	(1)	(2)	(3)	(4)	(5)
Urban	0.294***	0.274***	0.273***	0.197***	0.150***
	(0.025)	(0.025)	(0.026)	(0.036)	(0.030)
Female		−0.128***	−0.123***	0.005	
		(0.027)	(0.028)	(0.048)	
Years of education		−0.007	−0.008	0.021	
		(0.027)	(0.028)	(0.047)	
Years of education squared		0.001	0.001	0.000	
		(0.001)	(0.001)	(0.002)	
Normalized Ravens			0.033**	0.006	
			(0.016)	(0.028)	
Normalized Ravens squared			−0.011	0.017	
			(0.013)	(0.025)	
Individual fixed effects	N	N	N	N	Y
Time fixed effects	N	Y	Y	Y	Y
Switchers only				Y	
Number of observations	130322	130322	124481	38206	130322
Number of individuals	4718	4718	4452	1017	4718

Notes: Panel A uses data from the IFLS and Panel B uses data from the KLPS. Please refer to Section 3 for further details on the data and to the notes of Tables 4 and 5 for additional information on the variables. The dependent variable is log hours worked in wage and self-employment. If an individual has multiple jobs in the same time period, hours from all employment are included. The sample size in column 3 is smaller in Panel A because the Raven's test was administered only for a subset of the Indonesian sample. Column 4 only includes switchers, who are defined as individuals with at least one observation in both an urban and rural area. Each regression in columns 2–5 include quadratic controls for age. All regressions are clustered at the individual level. Robust standard errors are in parentheses, *** p<0.01, ** p<0.05, * p<0.1.

Table A10: Robustness to Alternative Agricultural Productivity Measures

(A) Indonesia

Definition of Agriculture	Productivity Measure Includes... Formal Wages	Self-Employed Profits	Dependent variable: Log Wage
Majority of hours in agriculture Main Estimation	✓	✓	0.077***
			(0.020)
Any hours in agriculture	✓	✓	0.040**
			(0.018)
All hours in agriculture	✓	✓	0.098***
			(0.019)
Majority of hours in agriculture	✓		-0.019
			(0.024)
Majority of hours in agriculture		✓	0.128***
			(0.030)

(B) Kenya

Definition of Agriculture	Productivity Measure Includes... Formal Wages	Self-Employed Profits	Dependent variable: Log Wage
Majority of hours in agriculture Main Estimation	✓	✓	0.014
			(0.106)
Any hours in agriculture	✓	✓	0.057
			(0.096)
All hours in agriculture	✓	✓	0.010
			(0.108)
Majority of hours in agriculture	✓		0.098
			(0.120)
Self-employment only		✓	0.031
			(0.177)

Notes: Panel A uses data from the IFLS and Panel B uses data from the KLPS. Each row shows the robustness results of a regression of log wages (calculated as earnings per hour) on a non-agricultural indicator, age squared, and time and individual fixed effects. In each panel, the estimate in row 1 can be found in Table 4, column 7; row 2 can be found in Appendix Table A6, column 4; row 3 in Appendix Table A6, column 8; row 4 in Appendix Table A12, column 4; and row 5 in Appendix Table A13, column 4. All regressions report standard errors clustered at the individual level. Robust standard errors are in parentheses, *** p<0.01, ** p<0.05, * p<0.1.

Table A11: Gap in Earnings for those Aged 30 or Younger, Indonesia

	Dependent variable: Log Earnings							
	(1)	(2)	(3)	(4) Log Wage	(5)	(6)	(7)	(8) Log Wage
Non-agricultural employment	0.531***	0.308***	0.160***	0.011				
	(0.020)	(0.018)	(0.030)	(0.037)				
Urban					0.399***	0.241***	0.077***	0.046**
					(0.015)	(0.013)	(0.019)	(0.023)
Individual fixed effects	N	N	Y	Y	N	N	Y	Y
Control variables and time FE	N	Y	Y	Y	N	Y	Y	Y
Number of observations	74817	74817	74817	74817	74817	74817	74817	74817
Number of individuals	18807	18807	18807	18807	18807	18807	18807	18807

Notes: This table uses data from the IFLS. The table repeats some of the analyses shown in Tables 4 and 5 but restricts the sample to observations where the individual is aged 30 years or fewer to allow better comparability to the KLPS sample. For columns 4 and 8, the dependent variable is the log of earnings divided by hours worked. Control variables include log hours, log hours squared, age, age squared, years of education, years of education squared and an indicator for being female. When also including individual fixed effects in columns 3, 4, 7 and 8, the control variables are reduced to only age squared. All regressions are clustered at the individual level. Robust standard errors are in parentheses, *** p<0.01, ** p<0.05, * p<0.1.

Table A12: Gap in Wage Earnings

(A) Indonesia

	Dependent variable: Log Wage Earnings							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
				Log Wage				Log Wage
Non-agricultural employment	1.035***	0.330***	0.154***	−0.019				
	(0.017)	(0.014)	(0.019)	(0.024)				
Urban					0.543***	0.178***	0.031**	0.009
					(0.015)	(0.010)	(0.013)	(0.015)
Individual fixed effects	N	N	Y	Y	N	N	Y	Y
Control variables and time FE	N	Y	Y	Y	N	Y	Y	Y
Number of observations	139846	139846	139846	139846	139846	139846	139846	139846
Number of individuals	22451	22451	22451	22451	22451	22451	22451	22451

(B) Kenya

	Dependent variable: Log Wage Earnings							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
				Log Wage				Log Wage
Non-agricultural employment	0.769***	0.548***	0.299***	0.098				
	(0.066)	(0.060)	(0.089)	(0.120)				
Urban					0.645***	0.553***	0.147***	0.087**
					(0.036)	(0.031)	(0.036)	(0.040)
Individual fixed effects	N	N	Y	Y	N	N	Y	Y
Control variables and time FE	N	Y	Y	Y	N	Y	Y	Y
Number of observations	94653	94653	94653	94653	94653	94653	94653	94653
Number of individuals	4020	4020	4020	4020	4020	4020	4020	4020

Notes: Panel A uses data from the IFLS, and Panel B uses data from the KLPS. The table repeats some of the analyses shown in Tables 4 and 5, but instead of using all available earnings as the dependent variable, this table only includes earnings from wage employment. For columns 4 and 8, the dependent variable is earnings from wage employment divided by hours worked. Control variables include log hours, log hours squared, age, age squared, years of education, years of education squared and an indicator for being female. When also including individual fixed effects in columns 3, 4, 7 and 8, the control variables are reduced to only age squared. All regressions are clustered at the individual level. Robust standard errors are in parentheses, *** p<0.01, ** p<0.05, * p<0.1.

Table A13: Gap in Self-Employment Earnings

(A) Indonesia

	Dependent variable: Log Self-Employment Earnings							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
				Log Wage				Log Wage
Non-agricultural employment	0.414***	0.311***	0.123***	0.128***				
	(0.017)	(0.016)	(0.026)	(0.030)				
Urban					0.393***	0.242***	0.042*	0.057**
					(0.017)	(0.015)	(0.022)	(0.026)
Individual fixed effects	N	N	Y	Y	N	N	Y	Y
Control variables and time FE	N	Y	Y	Y	N	Y	Y	Y
Number of observations	134153	134153	134153	134153	134153	134153	134153	134153
Number of individuals	17302	17302	17302	17302	17302	17302	17302	17302

(B) Kenya

	Dependent variable: Log Self-Employment Earnings							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
				Log Wage				Log Wage
Non-agricultural employment	0.451***	0.249**	−0.110	0.031				
	(0.114)	(0.099)	(0.162)	(0.177)				
Urban					0.598***	0.395***	0.050	0.053
					(0.082)	(0.076)	(0.128)	(0.151)
Individual fixed effects	N	N	Y	Y	N	N	Y	Y
Control variables and time FE	N	Y	Y	Y	N	Y	Y	Y
Number of observations	37064	37064	37064	37064	37064	37064	37064	37064
Number of individuals	1382	1382	1382	1382	1382	1382	1382	1382

Notes: Panel A uses data from the IFLS, and Panel B uses data from the KLPS. The table repeats some of the analyses shown in Tables 4 and 5, but instead of using all available earnings as the dependent variable, this table only includes earnings from self-employment. For columns 4 and 8, the dependent variable is earnings from self-employment divided by hours worked. Control variables include log hours, log hours squared, age, age squared, years of education, years of education squared and an indicator for being female. When also including individual fixed effects in columns 3, 4, 7 and 8, the control variables are reduced to only age squared. All regressions are clustered at the individual level. Robust standard errors are in parentheses, *** p<0.01, ** p<0.05, * p<0.1.

Table A14: Alternative Samples Kenya

(A) Subsistence agriculture included also if not main decision maker

	Dependent variable: Log Wage Earnings							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
				Log Wage				Log Wage
Non-agricultural employment	0.724***	0.471***	0.210**	0.018				
	(0.060)	(0.054)	(0.085)	(0.104)				
Urban					0.773***	0.601***	0.219***	0.123***
					(0.035)	(0.032)	(0.040)	(0.043)
Individual fixed effects	N	N	Y	Y	N	N	Y	Y
Control variables and time FE	N	Y	Y	Y	N	Y	Y	Y
Number of observations	130311	130311	130311	130311	130311	130311	130311	130311
Number of individuals	4717	4717	4717	4717	4717	4717	4717	4717

(B) Subsistence agriculture not included

	Dependent variable: Log Wage Earnings							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
				Log Wage				Log Wage
Non-agricultural employment	0.608***	0.394***	0.142	−0.014				
	(0.065)	(0.059)	(0.094)	(0.120)				
Urban					0.753***	0.594***	0.201***	0.108**
					(0.036)	(0.032)	(0.040)	(0.044)
Individual fixed effects	N	N	Y	Y	N	N	Y	Y
Control variables and time FE	N	Y	Y	Y	N	Y	Y	Y
Number of observations	128354	128354	128354	128354	128505	128354	128354	128354
Number of individuals	4611	4611	4611	4611	4624	4611	4611	4611

Notes: Panels A and B use data from the KLPS, described in Section 3. Panel A also includes productivity from subsistence agriculture if the individual is not the main decision maker for the agricultural activity. Panel B excludes all data from subsistence agriculture. For columns 4 and 8, the dependent variable is total earnings divided by hours worked. Control variables include log hours, log hours squared, age, age squared, years of education, years of education squared and an indicator for being female. When also including individual fixed effects in columns 3, 4, 7 and 8, the control variables are reduced to only age squared. All regressions are clustered at the individual level. Robust standard errors are in parentheses, *** p<0.01, ** p<0.05, * p<0.1.

Table A15: Unemployment and Job Search Behavior, Kenya

(A) Unemployment

	Dependent Variable: Unemployment or Subsistence Agriculture			Dependent Variable: Unemployment		
	(1)	(2)	(3)	(4)	(5)	(6)
Urban	−0.033***	−0.027**	−0.001	0.155***	0.141***	0.116***
	(0.012)	(0.012)	(0.021)	(0.008)	(0.008)	(0.012)
Individual fixed effects	N	N	Y	N	N	Y
Control variables and time FE	N	Y	Y	N	Y	Y
Mean dependent variable	0.297	0.297	0.297	0.080	0.080	0.080
Number of observations	10917	10917	10917	10917	10917	10917
Number of individuals	6794	6794	6794	6794	6794	6794

(B) Search Behavior

	Dependent variable: Total Hours Job Search		
	(1)	(2)	(3)
Urban	1.242***	1.216***	1.792***
	(0.144)	(0.150)	(0.266)
Individual fixed effects	N	N	Y
Control variables and time FE	N	Y	Y
Mean dependent variable	1.845	1.845	1.845
Number of observations	10917	10917	10917
Number of individuals	6794	6794	6794

Notes: Panel A reports urban gaps in unemployment. The first three columns define an individual as being unemployed if they are searching for work and have no income from wage or salary employment. The second three columns define an individual as being unemployed if they are searching for work and have no income from wage, salary, or proceeds from subsistence agriculture reported in the agricultural module. Sample sizes differ from analysis of wage gaps because questions about job search are contemporaneous to the time of the survey and are not retrospective. The dependent variable in Panel B is the number of hours a person reports to be searching for work; this variable equals 0 if the person is not searching for work. Like Panel A, data was only collected on search behavior contemporaneous to the time of the survey and thus sample sizes are smaller. Control variables include age, age squared, years of education, years of education squared and an indicator for being female. When also including individual fixed effects in columns 3 and 6, the control variables are reduced to only age squared. All regressions are clustered at the individual level. Robust standard errors are in parentheses, *** p<0.01, ** p<0.05, * p<0.1.

Table A16: Alternative Coefficient Standard Error Estimation

(A) Indonesia

	Dependent variable: Log Earnings							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
				Log Wage				Log Wage
Non-agricultural employment	0.715***	0.376***	0.239***	0.077***				
	(0.013)	(0.012)	(0.017)	(0.020)				
	[0.013]	[0.012]	[0.017]	[0.020]				
	⟦0.013⟧	⟦0.012⟧	⟦0.017⟧	⟦0.020⟧				
	{0.012}	{0.012}	{0.018}	{0.020}				
	⟨0.013⟩	⟨0.012⟩	⟨0.018⟩	⟨0.021⟩				
Urban					0.502***	0.225***	0.041***	0.033**
					(0.011)	(0.009)	(0.012)	(0.014)
					[0.011]	[0.009]	[0.012]	[0.014]
					⟦0.011⟧	⟦0.009⟧	⟦0.012⟧	⟦0.014⟧
					{0.012}	{0.009}	{0.013}	{0.014}
					⟨0.011⟩	⟨0.009⟩	⟨0.013⟩	⟨0.015⟩
Individual fixed effects	N	N	Y	Y	N	N	Y	Y
Control variables and time FE	N	Y	Y	Y	N	Y	Y	Y
Number of observations	258745	258745	258745	258745	258745	258745	258745	258745
Number of individuals	31537	31537	31537	31537	31537	31537	31537	31537

(B) Kenya

	Dependent variable: Log Earnings (in IDR)							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
				Log Wage				Log Wage
Non-agricultural employment	0.724***	0.470***	0.219**	0.014				
	(0.060)	(0.054)	(0.086)	(0.106)				
	[0.060]	[0.054]	[0.086]	[0.106]				
	⟦0.060⟧	⟦0.055⟧	⟦0.087⟧	⟦0.107⟧				
	{0.060}	{0.055}	{0.085}	{0.107}				
	⟨0.060⟩	⟨0.054⟩	⟨0.087⟩	⟨0.108⟩				
Urban					0.778***	0.604***	0.219***	0.123***
					(0.035)	(0.032)	(0.040)	(0.043)
					[0.035]	[0.032]	[0.040]	[0.043]
					⟦0.035⟧	⟦0.032⟧	⟦0.040⟧	⟦0.044⟧
					{0.035}	{0.032}	{0.040}	{0.044}
					⟨0.035⟩	⟨0.032⟩	⟨0.041⟩	⟨0.044⟩
Individual fixed effects	N	N	Y	Y	N	N	Y	Y
Control variables and time FE	N	Y	Y	Y	N	Y	Y	Y
Number of observations	130322	130322	130322	130322	130322	130322	130322	130322
Number of individuals	4718	4718	4718	4718	4718	4718	4718	4718

Notes: Panel A uses data from the IFLS, and Panel B uses data from the KLPS. The table repeats some of the analyses shown in Tables 4 and 5 and presents cluster robust standard errors computed several ways. For each coefficient, standard errors in parentheses in row 2 are the default standard errors reported by Stata. Rows 3 and 4 in single and double square brackets, respectively, report cluster robust standard errors CR2 and CR3 (Bell and McCaffrey 2002) that correct variance matrix bias by transforming residuals (see also Cameron and Miller, 2015). Row 5 in curly braces reports block bootstrapped errors for 1,000 bootstrap samples between stars. And, Row 6 in triangular brackets reports standard errors with Young (2016) effective degrees of freedom corrections. For columns 4 and 8, the dependent variable is the log of earnings divided by hours worked. Control variables include log hours, log hours squared, age, age squared, years of education, years of education squared and an indicator for being female. When also including individual fixed effects in columns 3, 4, 7 and 8, the control variables are reduced to only age squared. Significance stars reported reflect hypothesis tests using t-statistics computed from default standard errors, *** p<0.01, ** p<0.05, * p<0.1.

Table A17: Integenerational Correlations of Cognitive Measures

(A) Indonesia

	Dependent variable: Normalized Ravens				
	(1)	(2)	(3)	(4)	(5)
Child Covariates:					
Born Urban	0.189***	0.190***	0.191***	0.150***	0.141***
	(0.040)	(0.040)	(0.045)	(0.045)	(0.045)
Female		−0.154***	−0.157***	−0.162***	−0.159***
		(0.039)	(0.040)	(0.038)	(0.038)
Parent (Averaged) Covariates:					
Born Urban			−0.005	−0.076	−0.087*
			(0.052)	(0.051)	(0.051)
Age at Birth			−0.001	−0.001	−0.000
			(0.001)	(0.001)	(0.002)
Years of Education				0.036***	0.074***
				(0.006)	(0.024)
Normalized Ravens				0.202***	0.208***
				(0.026)	(0.026)
Age, Education, and Ravens Squared	N	N	N	N	Y
Number of observations	1725	1725	1725	1725	1725

(B) Kenya

	Dependent variable: Normalized Cognitive Ability Index				
	(1)	(2)	(3)	(4)	(5)
Child Covariates:					
Born Urban	0.344***	0.345***	0.369***	0.258***	0.258***
	(0.082)	(0.082)	(0.083)	(0.085)	(0.085)
Female		0.111	0.111	0.102	0.108
		(0.070)	(0.070)	(0.069)	(0.069)
KLPS Parent Covariates:					
Female			0.271***	0.308***	0.309***
			(0.075)	(0.073)	(0.074)
Age at Birth			0.002	0.013	0.270
			(0.014)	(0.014)	(0.195)
Years of Education				0.065***	0.068
				(0.014)	(0.067)
Normalized Ravens				0.055	0.053
				(0.040)	(0.041)
Age, Education, and Ravens Squared	N	N	N	N	Y
Number of observations	864	864	864	864	864

Notes: Notes: Panel A uses data from individuals in the IFLS, and Panel B uses data from a sample of children aged 3–5 of the KLPS adults. The Cognitive Ability index in Panel B is a composite of z-scores from six different tests of language, attention, memory, perception, and fine motor skills. Ravens matrices scores and the Cognitive Ability index are normalized to have mean zero and unit variance for full-year and six-month child age bins, respectively. In Panel A, parent covariates are averaged when both parents are available. In Panel B, only the covariates for the adult KLPS respondent are available (not for the spouse of the adult KLPS respondent). Regressions are clustered at the individual level in Panel A and at the parent level in Panel B. Robust standard errors are in parentheses, *** p<0.01, ** p<0.05, * p<0.1.

Table A18: Correlates of Meals Eaten—Kenya

	(1) Log Consumption	(2) Log Earnings	(3) Log Wage
Log(Meals)	0.194*	0.278***	0.228***
	(0.090)	(0.065)	(0.066)
Number of observations	1062	4693	4315

Notes: Each cell reports a regression coefficient with the log of meals as the independent variable; dependent variables are listed in the header of the table. These regressions do not have the sample restrictions found in Table 2. Log of household per capita consumption in column 1 is available only for a subset of individuals from KLPS 3. Robust standard errors clustered by individual reported below in parentheses, *** p<0.01, ** p<0.05, * p<0.1.

Table A19: Gaps in Consumption

(A) Indonesia

	Dependent variable: Log Consumption					
	(1)	(2)	(3)	(4)	(5)	(6)
Non-agricultural employment	0.441***	0.223***	0.076***			
	(0.007)	(0.006)	(0.010)			
Urban				0.379***	0.183***	0.050***
				(0.006)	(0.006)	(0.009)
Individual fixed effects	N	N	Y	N	N	Y
Control variables and time FE	N	Y	Y	N	Y	Y
Number of observations	77303	77303	77303	77303	77303	77303
Number of individuals	34143	34143	34143	34143	34143	34143

(B) Kenya

	Dependent variable: Log Meals Eaten					
	(1)	(2)	(3)	(4)	(5)	(6)
Non-agricultural employment	0.078***	0.059***	−0.090*			
	(0.016)	(0.017)	(0.049)			
Urban				0.029***	0.030***	−0.023
				(0.010)	(0.011)	(0.040)
Individual fixed effects	N	N	Y	N	N	Y
Control variables and time FE	N	Y	Y	N	Y	Y
Number of observations	4203	4203	4203	4203	4203	4203
Number of individuals	3601	3601	3601	3601	3601	3601

Notes: Panel A uses data on total consumption from the IFLS, and Panel B uses data on meals eaten in the last day from the KLPS. Unlike previous tables, the sample includes individuals with and without earnings measures. Consumption data in the IFLS are obtained by adding up the value of food and non-food consumption in Indonesian Rupiah at the household level and dividing this by the number of household members. The data was collected for each of the five waves so each household has five observations at most. Separate analyses by food and non-food consumption in Indonesia can be found in Appendix Table A19, and Appendix Table A20 provides consumption analyses when using the sample with positive earnings measures. Data on meals eaten in Kenya are available from KLPS rounds 2 and 3 and refer to the day prior to the survey date. In the analysis sample, 0.6% of individual-time observations ate no meals in the prior day, 10.9% ate one meal, 53.2% ate two meals, 34.0% ate three meals, and 1.3% ate four or more. Control variables in both panels include age, age squared, years of education, years of education squared, and an indicator for being female. When also including individual fixed effects in columns 3 and 6, the control variables are reduced to only age squared because the others are absorbed by the individual fixed effects. All regressions are clustered at the individual level. Robust standard errors are in parentheses, *** p<0.01, ** p<0.05, * p<0.1.

Table A20: Gap in Food and Non-Food Consumption, Indonesia

(A) Food Consumption

	Dependent variable: Log Food Consumption					
	(1)	(2)	(3)	(4)	(5)	(6)
Non-agricultural employment	0.275***	0.131***	0.058***			
	(0.006)	(0.006)	(0.010)			
Urban				0.252***	0.122***	0.038***
				(0.006)	(0.005)	(0.009)
Individual fixed effects	N	N	Y	N	N	Y
Control variables and time FE	N	Y	Y	N	Y	Y
Number of observations	77303	77303	77303	77303	77303	77303
Number of individuals	34143	34143	34143	34143	34143	34143

(B) Non-Food Consumption

	Dependent variable: Log Non-Food Consumption					
	(1)	(2)	(3)	(4)	(5)	(6)
Non-agricultural employment	0.728***	0.399***	0.104***			
	(0.009)	(0.009)	(0.015)			
Urban				0.593***	0.299***	0.060***
				(0.009)	(0.008)	(0.013)
Individual fixed effects	N	N	Y	N	N	Y
Control variables and time FE	N	Y	Y	N	Y	Y
Number of observations	77303	77303	77303	77303	77303	77303
Number of individuals	34143	34143	34143	34143	34143	34143

Notes: Both panels use data from the IFLS. Panels A and B repeat the consumption analyses shown in Appendix Table A18, broken down by food and non-food consumption respectively. Please refer to Appendix Table A18 for further details. Control variables include age, age squared, years of education, years of education squared and an indicator for being female. When also including individual fixed effects in columns 3 and 6, the control variables are reduced to only age squared. All regressions are clustered at the individual level. Robust standard errors are in parentheses, *** p<0.01, ** p<0.05, * p<0.1.

Table A21: Gap in Consumption (Main Analysis Sample), Indonesia

(A) Total Consumption

	Dependent variable: Log Consumption					
	(1)	(2)	(3)	(4)	(5)	(6)
Non-agricultural employment	0.397***	0.175***	0.058***			
	(0.007)	(0.007)	(0.012)			
Urban				0.278***	0.089***	−0.033***
				(0.007)	(0.006)	(0.010)
Individual fixed effects	N	N	Y	N	N	Y
Control variables and time FE	N	Y	Y	N	Y	Y
Number of observations	64685	64685	64685	64685	64685	64685
Number of individuals	30173	30173	30173	30173	30173	30173

(B) Food Consumption

	Dependent variable: Log Food Consumption					
	(1)	(2)	(3)	(4)	(5)	(6)
Non-agricultural employment	0.233***	0.086***	0.040***			
	(0.007)	(0.007)	(0.011)			
Urban				0.155***	0.029***	−0.052***
				(0.006)	(0.006)	(0.009)
Individual fixed effects	N	N	Y	N	N	Y
Control variables and time FE	N	Y	Y	N	Y	Y
Number of observations	64685	64685	64685	64685	64685	64685
Number of individuals	30173	30173	30173	30173	30173	30173

(C) Non-Food Consumption

	Dependent variable: Log Non-Food Consumption					
	(1)	(2)	(3)	(4)	(5)	(6)
Non-agricultural employment	0.679***	0.341***	0.086***			
	(0.011)	(0.010)	(0.017)			
Urban				0.482***	0.196***	−0.017
				(0.010)	(0.009)	(0.014)
Individual fixed effects	N	N	Y	N	N	Y
Control variables and time FE	N	Y	Y	N	Y	Y
Number of observations	64685	64685	64685	64685	64685	64685
Number of individuals	30173	30173	30173	30173	30173	30173

Notes: All regressions use data from the IFLS. This table repeats the analyses shown in Appendix Table A18 and A11 using the main analysis sample, which excludes individual-year observations without earnings measures. Thus, the sample size is smaller than in Appendix Table A18. Control variables include age, age squared, years of education, years of education squared and an indicator for being female. When also including individual fixed effects in columns 3 and 6, the control variables are reduced to only age squared. All regressions are clustered at the individual level. Robust standard errors are in parentheses, *** p<0.01, ** p<0.05, * p<0.1.

Table A22: Gap in Consumption for those Born in Rural and Urban Areas, Indonesia

(A) Indonesian individuals born in rural areas (Dependent variable: Log Consumption)

	Full Consumption Sample			Main Analysis Sample		
	(1)	(2)	(3)	(4)	(5)	(6)
Urban	0.351***	0.184***	0.049***	0.265***	0.099***	−0.033***
	(0.008)	(0.007)	(0.010)	(0.009)	(0.008)	(0.012)
Individual fixed effects	N	N	Y	N	N	Y
Control variables and time FE	N	Y	Y	N	Y	Y
Number of observations	56248	56248	56248	47126	47126	47126
Number of individuals	23857	23857	23857	21069	21069	21069

(B) Indonesian individuals born in urban areas (Dependent variable: Log Consumption)

	Full Consumption Sample			Main Analysis Sample		
	(1)	(2)	(3)	(4)	(5)	(6)
Urban	0.252***	0.118***	0.037**	0.154***	0.022	−0.030
	(0.015)	(0.013)	(0.019)	(0.017)	(0.014)	(0.020)
Individual fixed effects	N	N	Y	N	N	Y
Control variables and time FE	N	Y	Y	N	Y	Y
Number of observations	20864	20864	20864	18653	18653	18653
Number of individuals	10167	10167	10167	9277	9277	9277

Notes: Both panels use data from the IFLS. Panels A and B repeat the consumption analyses shown in Appendix Table A18, broken down by those born in rural and urban areas respectively. Please refer to Appendix Table A18 for further details. Control variables include age, age squared, years of education, years of education squared and an indicator for being female. When also including individual fixed effects in columns 3 and 6, the control variables are reduced to only age squared. All regressions are clustered at the individual level. Robust standard errors are in parentheses, *** p<0.01, ** p<0.05, * p<0.1.

Table A23: Urban/Rural Gap in Wages for Top 5 Cities

(A) Indonesia

	Dependent variable: Log Wages			
	(1)	(2)	(3)	(4)
Urban	0.345***	0.291***	0.074***	0.032**
	(0.011)	(0.012)	(0.011)	(0.014)
Jakarta (population 10 million)		0.293***	0.315***	0.025
		(0.020)	(0.017)	(0.038)
Surabaya (population 2.8 million)		−0.012	−0.003	0.012
		(0.056)	(0.047)	(0.110)
Bandung (population 2.6 million)		0.239***	0.153***	0.094
		(0.060)	(0.047)	(0.110)
Medan (population 2.5 million)		0.286***	0.269***	−0.022
		(0.048)	(0.045)	(0.139)
Bekasi (population 2.5 million)		0.682***	0.477***	0.151*
		(0.060)	(0.055)	(0.086)
Individual fixed effects	N	N	N	Y
Control variables and time FE	N	N	Y	Y
Number of observations	258745	258745	258745	258745
Number of individuals	31537	31537	31537	31537

(B) Kenya

	Dependent variable: Log Wages			
	(1)	(2)	(3)	(4)
Urban	0.484***	0.309***	0.272***	0.048
	(0.036)	(0.055)	(0.050)	(0.060)
Nairobi (population 3.4 million)		0.280***	0.262***	0.139**
		(0.056)	(0.050)	(0.058)
Mombasa (population 1.2 million)		0.274***	0.262***	0.263***
		(0.074)	(0.069)	(0.088)
Kisumu (population 0.4 million)		−0.065	−0.006	−0.140
		(0.127)	(0.119)	(0.106)
Nakuru (population 0.3 million)		0.232**	0.156*	0.201
		(0.109)	(0.089)	(0.149)
Eldoret (population 0.3 million)		0.066	0.026	−0.221*
		(0.143)	(0.146)	(0.127)
Individual fixed effects	N	N	N	Y
Control variables and time FE	N	N	Y	Y
Number of observations	130322	130322	130322	130322
Number of individuals	4718	4718	4718	4718

Notes: Panel A uses data from the IFLS and Panel B uses data from the KLPS. Please refer to Section 3 for further details on the data and to the notes of Table 5 for additional information on the variables. The covariate "Urban" is an indicator variable that equals 1 if the person lives in an urban area, and five city indicators are included for the five most populous cities in Indonesia and Kenya, respectively. Control variables include age, age squared, years of education, years of education squared and an indicator for being female. When also including individual fixed effects in columns 4, the control variables are reduced to only age squared. All regressions are clustered at the individual level. Robust standard errors are in parentheses, *** p<0.01, ** p<0.05, * p<0.1.

Appendix B: Sector-specific Productivity — Absolute and Comparative

In the conceptual framework, the richest, most flexible specification of human capital allows for individual sector-specific productivity $\theta_{is} = \boldsymbol{x_i'\beta_s} + \boldsymbol{\eta_{is}}$. Analysis of these productivities has been given renewed focus in Lagakos and Waugh (2013).

Utilizing panel data, we estimate a modified version of equation 4 replacing the individual fixed effect with an individual-sector fixed effect.[1] We recover these estimates, and then normalize the mean of the fixed effects of permanent rural residents (non-movers) to be zero. Appendix Figure A8 presents the joint distributions of these estimated individual productivities by sector. Panel A includes Indonesians born in rural areas. It is apparent that rural-to-urban migrants are positively selected relative to non-migrants, with an average rural wage approximately 18 log points higher than non-migrants. These individuals experience only a 2 log point average increase in their wage upon migration to an urban area. Panel B presents the same exercise with Indonesians born in urban areas. Here, there appears to be little selection into rural migration, with the average mover having approximately 3 log points higher wages when still in urban areas, and no change in rural wages among movers.

Panel C presents results in Kenya (all of whom were rural residents as children) that are analogous to panel A. Compared to Indonesia, there appears to be even more positive selection among urban migrants in Kenya (at 39 log points) as well as a moderate positive urban premium of roughly 12 log points, which is nearly identical to the regression adjusted estimate presented above.

---

[1] This procedure is similar in spirit to the correlated random coefficient models utilized to analyze heterogeneous returns to hybrid seed adoption (Suri 2011) and labor unions' effects on wages (Card 1996, Lemieux 1998), although our approach makes fewer assumptions and is meant to be more descriptive.

Note that the realizations of roughly half of migrants fall below the 45 degree line in the three panels of Appendix Figure A8, which taken literally means that they experience higher earnings in rural than urban areas. This is consistent with the empirical finding of zero or small positive sectoral productivity gaps.

This exercise is meant to be descriptive, and we interpret the relationships between the estimated individual urban and rural productivities with some caution, in part because the estimates are subject to measurement error and thus the fitted regression line may experience attenuation bias. With these caveats in mind, note that all three plots appear to show that absolute advantage plays a role in wage determination: individuals who have high rural productivity tend to have high urban productivity, and vice versa, indicated by the positive slope.

Appendix C: Productivity versus Living Standards

Section 4 establishes a reduction of roughly 90% in estimated sectoral productivity gaps once individual fixed effects are included in the analysis (Figure 1). While wages are closely associated to labor productivity measures, productivity and "utility" may diverge for many reasons, including price differences across regions, amenities, unemployment, and other factors. For instance, there could be considerable individual heterogeneity in the taste for rural versus urban amenities, e.g., comforts of home, ethnic homogeneity, better informal insurance, etc., in rural areas versus cosmopolitan cities' better public goods and more novelty (but downsides too, such as crime). Moving itself may also impose large utility costs (Kleemans 2016).

Although it is impossible to fully capture these factors and convincingly measure individual welfare, to get somewhat closer to differences in living standards, we draw on detailed consumption data from Indonesia. In the main specifications, we are able to include all individuals who have such consumption data, even if they lack earnings measures. Consumption expenditures may also more accurately capture total household income in low-income settings like ours with extensive home production, informal employment, unemployment and underemployment.

In the Indonesia Family Life Survey (IFLS) consumption data were collected by directly asking households the value in Indonesian Rupiah of all food and non-food purchases and consumption in the last month, similar to consumption data collection in the World Bank's Living Standards Measurement Surveys.[2] In contrast to the retrospective earnings data in the

[2] Note that for a small number of frequently-consumed items, information was collected for the last week, and for a few low-frequency items, data was collected for the last year.

IFLS, the consumption data are all contemporaneous to the survey. Consumption data were collected at the household level, which we divide by the number of household members to obtain a per capita measure, and are presented in real terms, taking into account prices in rural and urban areas (see footnote in the main text for more information on this urban price deflator). The consumption sample includes 77,303 individual-year observations from 34,143 individuals in IFLS rounds 1–5. For the consumption analysis, we expand the sample to also include individuals without current earnings data; we also perform a robustness check using the main productivity sample.

Detailed consumption expenditure data for Kenya were collected for a subset of individuals in the third round of the Kenya Life Panel Survey. However, because it was only collected for one round, we are unable to utilize it in panel estimation. Instead, in the panel analysis we utilize a proxy for consumption, the number of meals eaten in the previous day, which is available in both KLPS Rounds 2 and 3. Reassuringly, meals eaten is strongly correlated with our primary measures of labor productivity as well as consumption expenditures per capita (in KLPS-3); see Appendix Table A18. As with Indonesia, in the meal consumption analysis, we are able to expand the sample to also include individuals without current earnings data.

Appendix Table A19 presents the main results of the consumption gap between non-agriculture and agriculture. The initial gap is large and similar the productivity gap at 44 log points (Panel A). The gap falls considerably when including time fixed effects and control variables in column 2, and falls to only 7.6 log points when also including individual fixed

effects in column 3. A similar pattern is presented for the urban-rural consumption gap in columns 4, 5, and 6: the gap declines from 38 log points to 5 log points.[3]

We next explore the estimated urban consumption premium for those born in rural versus urban areas (Appendix Table A22). In the preferred specification with individual fixed effects (col. 3 for the full consumption sample, col. 6 for the sample included in the main analysis), the urban consumption premium is larger for those born in rural areas (4.9 log points or -3.3 log points, respectively, Panel A), than those born in urban areas (3.7 log points or -3.0 log points, respectively, Panel B), and the difference is highly significant (p-value<0.001). As with the earnings results, this is consistent with the predictions of the selection model (in Section 2) and suggests the urban premium is bounded rather tightly around zero.

The consumption proxy measure in the KLPS tells a similar story. The raw gap in meals eaten in Kenya between those in non-agriculture versus agricultural employment is positive and statistically significant, though smaller than the earnings gap (Appendix Table A19, Panel B); differences in magnitude are difficult to interpret given the different nature of the meals measure, and the possibility that it changes most at very low levels of income. Mirroring the broad pattern observed for labor productivity, this gap falls when including controls, and is actually slightly negative when including individual fixed effects (columns 1-3); a similar pattern holds for the urban-rural gap (columns 4-6).

---

[3] Appendix Table A20 shows the gap in both food and non-food consumption (Panels A and B, respectively). The gaps in both components of consumption see reductions of 79-90% when including individual fixed effects. Appendix Table A21 repeats the consumption analyses on the main analysis sample (i.e., those with earnings data) for total consumption (Panel A) and by food and non-food consumption (Panels B and C, respectively), and results are similar.

Appendix D: Productivity Gaps in Big Cities

Section 4 shows large reductions in rural-urban productivity gaps once individual characteristics and fixed effects are controlled for, and also shows that larger gaps do not emerge over time. In this appendix, we explore whether productivity gaps are larger for large cities. To examine this, we first repeat the main urban productivity gap analysis (from Table 5) and then include a breakdown into the five highest population cities in each country, in Appendix Table A23. In Indonesia, all five cities are larger than 2 million inhabitants, with the capital Jakarta at 10 million. Kenya's capital Nairobi has 3.4 million people, the second largest city (Mombasa) has nearly one million, while the other three cities in Kenya are smaller. The capitals are also the largest destinations for urban migrants in each country.

There is mixed evidence on the extent of big city productivity effects. There is no evidence for significantly larger effects in Indonesia's largest cities, Jakarta, although we see slightly larger productivity gaps for the city of Bekasi (column 4 of Panel A). There is some evidence of significant positive urban productivity gains in the two largest Kenyan cities, Nairobi and Mombasa (Panel B). The total urban effect is moderate and statistically significant in the capital of Nairobi, at 14 log points.[4]

[4] While this analysis finds mixed evidence of an overall big city effect in Indonesia and Kenya, we also assess whether effects might manifest over a longer time horizon by repeating the event study analysis over a five year time horizon separately for Jakarta and Nairobi. These figures show no clear evidence of differentially positive dynamic effects in capital cities: differences with other cities are imprecisely estimated and generally not significant (not shown).