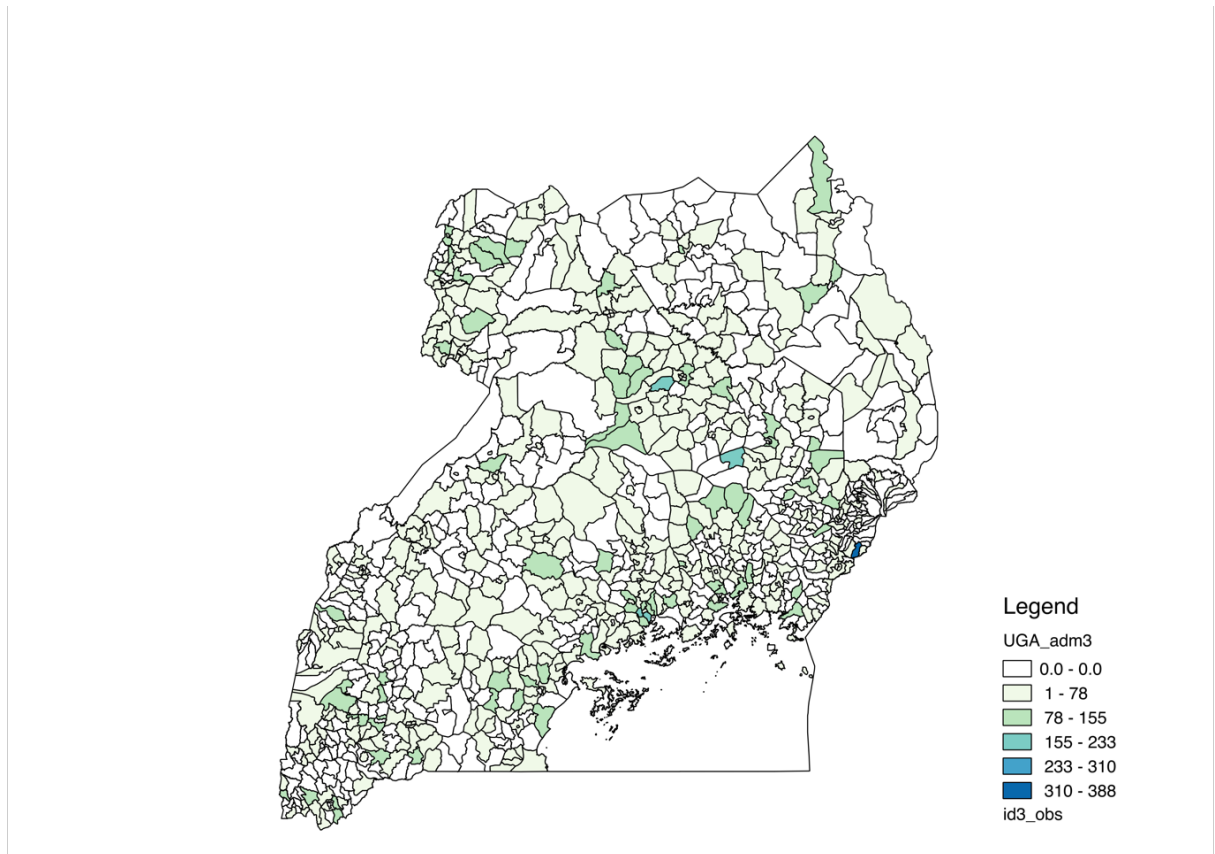


ONLINE APPENDIX - Not for publication

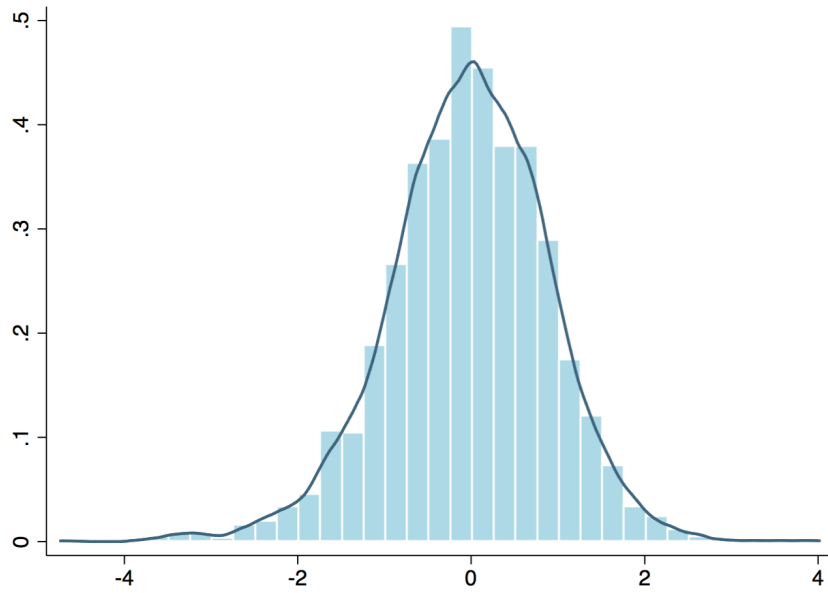
A Additional figures and tables

Figure A.1: Sample coverage



Notes: Figure depicts the number of observations per county.

Figure A.2: Distribution of farm productivity ($\ln s_i$)



Notes: The estimated production function parameters are $\hat{\alpha} = 0.526$ and $\hat{\gamma} = 0.708$. The difference between the 90th and 10th percentile is 2.23.

Figure A.3: Yields ($\ln Y/T$) and farm productivity ($\ln s_i$)

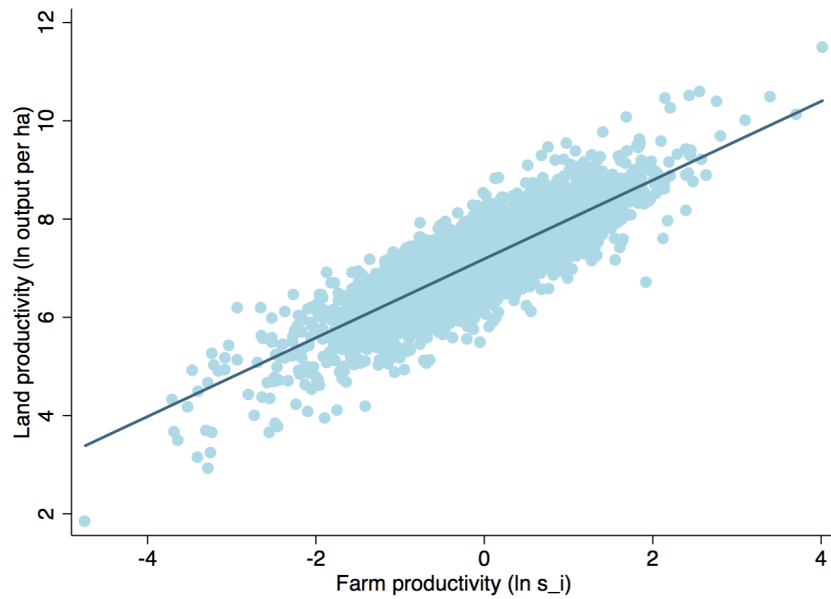
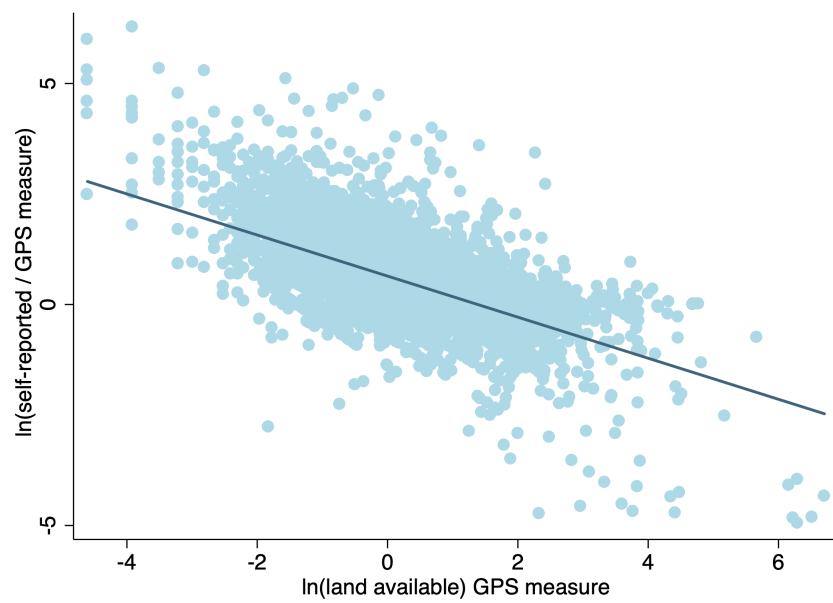


Figure A.4: Systematic measurement error in self-reported available land



Notes: Vertical axis is a proxy of measurement error = log of ratio of self-reported to GPS measure of available land.

Table A.1: Production function estimates

	ln(output)						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
ln(land)	0.372*** (0.020)	0.341*** (0.020)	0.355*** (0.020)	0.392*** (0.071)	0.048** (0.020)	0.298*** (0.021)	0.459*** (0.047)
ln(total labor)	0.336*** (0.017)	0.339*** (0.019)			0.428*** (0.021)	0.387*** (0.025)	0.418*** (0.077)
ln(domestic labor)			0.237*** (0.017)	0.296** (0.149)			
ln(hired labor)			0.132*** (0.011)	0.131*** (0.011)			
ln(value of tools and machinery)						0.041** (0.018)	
Method	Baseline	Baseline + agric. pract.	Baseline + disagg. labor	IV	Baseline + GPS measure	Baseline + capital	First Diff.
Implied γ	0.708	0.681	0.724	0.819	0.476	0.725	0.877
Implied α	0.526	0.502	0.490	0.479	0.101	0.533	0.476
Observations	15,541	14,361	14,361	13,933	10,789	11,535	8,082
No. farmers	3,457	3,403	3,403	3,356	2,617	3,321	2,118
R-squared	0.154	0.155	0.155		0.120	0.187	

Notes: Robust standard errors in parentheses. Standard errors are clustered at household level. * denotes significant at 10%, ** significant at 5% and *** significant at 1%. All regressions include region-by-period fixed effects and weather controls. Columns 1 to 6 also include household fixed effects. Columns 2 to 5 include indicators of agricultural practices (fertilizers, pesticides, improved seeds, intercropping, hired labor, and tenure of bulls/oxen). Column 3 disaggregates measure of total labor into domestic and hired labor. Column 4 uses land available and no. of household members who work in farm in last year as instruments for land cultivated and domestic labor. All regressions use land cultivated as measure of land, except for column 5 which uses GPS measure of land available. Column 6 adds the value of farm implements and machinery as a continuous measure of capital. Column 7 replicates the first difference panel model suggested in Shenoy (2017). Land measured in has. Labor measured in person-days.

Table A.2: Main results using available land as measure of size

	ln(output/land cultivated)				Farm productivity	
	(1)	(2)	(3)	(4)	(5)	(6)
ln(land available)	-0.073*** (0.013)	-0.101*** (0.013)	-0.230*** (0.022)	-0.037* (0.020)	0.188*** (0.011)	0.251*** (0.019)
Controls	No	Yes	Yes	Yes	Yes	Yes
Household FE	No	No	Yes	No	No	No
No. obs.	16,010	14,532	15,740	3,252	16,373	3,249
R-squared	0.003	0.153	0.057	0.250	0.392	0.350

Notes: Robust standard errors in parentheses. Standard errors are clustered at household level. * denotes significant at 10%, ** significant at 5% and *** significant at 1%. All regressions (except column 1) include soil and farmer controls similar to Table 2, as well as district fixed effects. Columns 2 to 4 also includes region-by-period fixed effects, while column 3 adds household fixed effects. Columns 4 and 6 use a cross-section of farmers obtained by collapsing the panel data at household level taking a simple average.

Table A.3: Yields and farm size using sub-sample of specialized farmers

	Outcome variable: $\ln(\text{output}/\text{land cultivated})$			
	(1)	(2)	(3)	(4)
$\ln(\text{land cultivated})$	-0.167** (0.068)	-0.317*** (0.050)	-0.164*** (0.050)	-0.292*** (0.056)
Sample:	Monocrop farmers	50% or more of farm land planted with: cassava maize beans		
No. obs.	870	1,366	867	673
R-squared	0.266	0.281	0.334	0.446

Notes: Robust standard errors in parentheses. Standard errors are clustered at household level. * denotes significant at 10%, ** significant at 5% and *** significant at 1%. Table replicates the results in column 2 of Table 2 using a sub-sample of farmers that planted a single crop (column 1) or that planted more than 50% of their farm land with one of the major crops: cassava, maize or beans (columns 2 to 4).

Table A.4: Farm productivity and farm size using sub-sample of specialized farmers

	Outcome variable = farm productivity			
	(1)	(2)	(3)	(4)
$\ln(\text{land cultivated})$	0.238*** (0.036)	0.175*** (0.030)	0.247*** (0.031)	0.209*** (0.034)
Sample:	Monocrop farmers	50% or more of farm land planted with: cassava maize beans		
No. obs.	870	1,365	866	673
R-squared	0.536	0.461	0.618	0.647

Notes: Robust standard errors in parentheses. Standard errors are clustered at household level. * denotes significant at 10%, ** significant at 5% and *** significant at 1%. Table replicates the results in column 1 of Table 3 using a sub-sample of farmers that planted a single crop (column 1) or that planted more than 50% of their farm land with one of the major crops: cassava, maize or beans (columns 2 to 4).

Table A.5: Yields and farm size using regional and local prices

	Outcome variable: $\ln(\text{output}/\text{land cultivated})$			
	(1)	(2)	(3)	(4)
$\ln(\text{land cultivated})$	-0.160*** (0.018)		-0.282*** (0.025)	
$\ln(\text{land available})$ GPS		-0.583*** (0.019)		-0.637*** (0.023)
Output prices	Regional (n=5)		District (n=109)	
No. obs.	14,685	10,330	7,582	5,601
R-squared	0.235	0.365	0.276	0.422

Notes: Robust standard errors in parentheses. Standard errors are clustered at household level. * denotes significant at 10%, ** significant at 5% and *** significant at 1%. All regressions include soil, farmer and weather controls similar to column 2 of Table 2, as well as district fixed effects. Columns 1 to 2 calculate real agricultural output (at 2009 prices) using median prices by region, while columns 3 and 4 use median prices by district.

Table A.6: Farm productivity and farm size using regional and local prices

	Outcome variable = farm productivity			
	(1)	(2)	(3)	(4)
$\ln(\text{land cultivated})$	0.226*** (0.014)		0.177*** (0.020)	
$\ln(\text{land available})$ GPS		0.157*** (0.013)		0.142*** (0.019)
Output prices	Regional (n=5)		District (n=109)	
No. obs.	15,368	11,146	13,640	9,986
R-squared	0.442	0.495	0.465	0.478

Notes: Robust standard errors in parentheses. Standard errors are clustered at household level. * denotes significant at 10%, ** significant at 5% and *** significant at 1%. All regressions include soil and farmer controls similar to column 1 of Table 3, as well as district fixed effects. Columns 1 to 2 calculate real agricultural output (at 2009 prices) using median prices by region, while columns 3 and 4 use median prices by district.

Table A.7: Farm productivity and farm size using alternative TFP estimates

	Farm productivity ($\ln s_i$)		
	(1)	(2)	(3)
$\ln(\text{land cultivated})$	0.198*** (0.011)	0.148*** (0.012)	0.478*** (0.017)
Prod. function used used to estimate s_i	Baseline	Baseline + capital	First Diff.
No. obs.	15,363	11,184	12,677
R-squared	0.399	0.759	0.395

Notes: Robust standard errors in parentheses. Standard errors are clustered at household level. * denotes significant at 10%, ** significant at 5% and *** significant at 1%. All regressions include soil and farmer controls and district fixed effects similar to column 1 of Table 3. Columns 1 uses the baseline estimates of TFP, while columns 2 and 3 use alternative estimates from columns 6 and 7 of Table A.1. These estimates are obtained from a model with a measure of capital (tools and machinery) and a first difference panel model suggested by Shenoy (2017).

Table A.8: Main crops

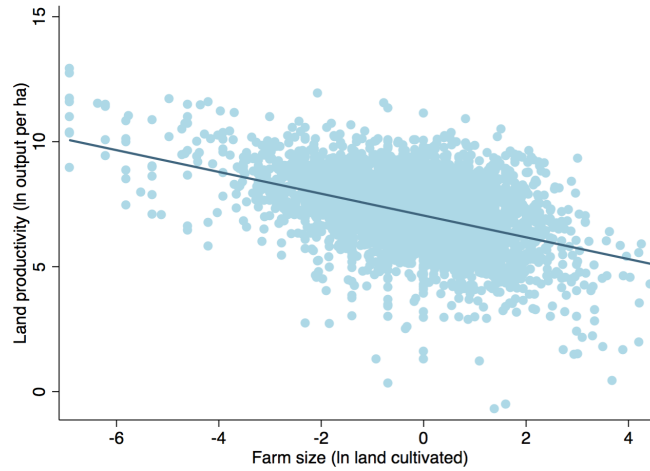
Crops	Average % area planted	% farmers who plant crop
Cassava	17.2	58.4
Beans	14.7	60.7
Maize	14.2	55.0
Banana food	13.8	47.5
Sweet potatoes	8.7	39.5
Ground nuts	4.2	20.7
Sorghum	4.0	13.2
Coffee	3.7	19.5
Millet	2.9	12.0

Table A.9: Usage of farm implements and machinery

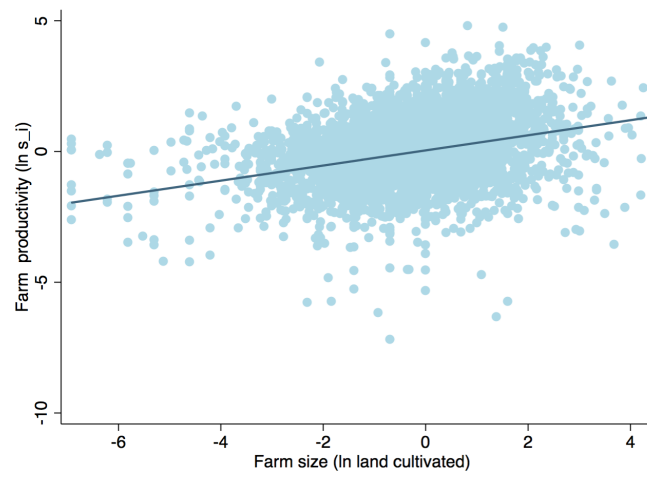
Farm implement or machinery	% farmers use tool
Hoe	99.4
Ploughs	3.8
Pangas (machetes)	86.5
Slashers	30.4
Wheel barrows	10.1
Tractor	0.3
Watering cans	4.1
Pruning knives	11.1
Pruning saws	0.6
Chain/band saws	0.5
Sheller	0.3
Spade	21.9
Fork hoe	6.7
Ox-plough	7.4
Trailer	0.1
Harrow/cultivator	0.4
Weeder	0.2
Planter	0.1
Sprayer	14.7
Pail	2.0
Milk can	1.2

B Evidence from other countries

Figure B.1: Farm size and productivity - Peru

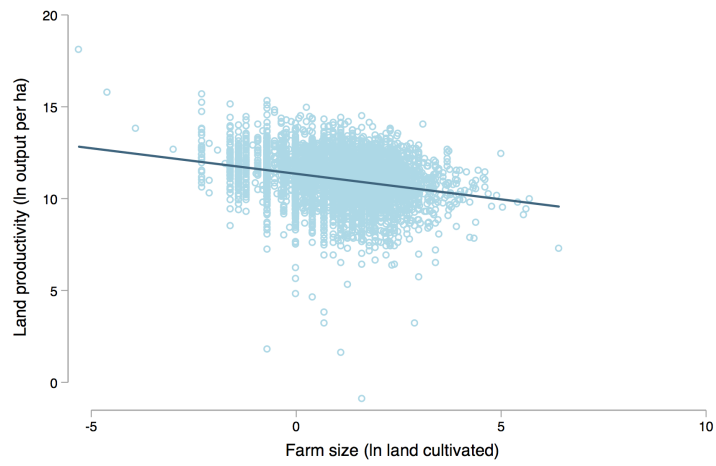


(a) Land productivity ($\ln Y/T$)

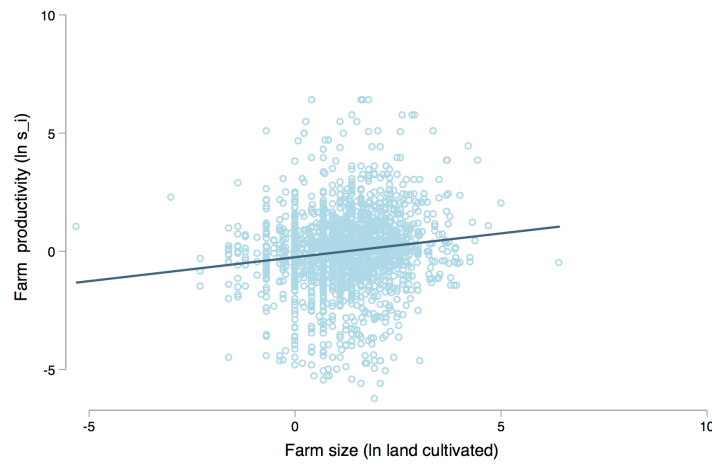


(b) Farm productivity ($\ln s_i$)

Figure B.2: Farm size and productivity - Tanzania

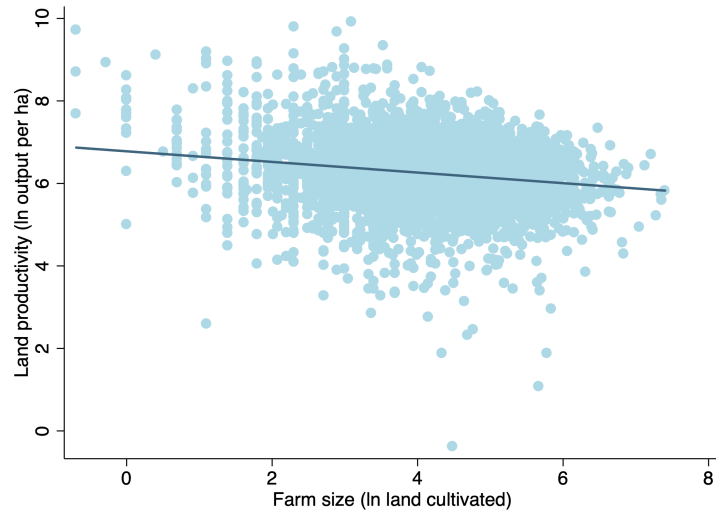


(a) Land productivity ($\ln(Y/T)$)

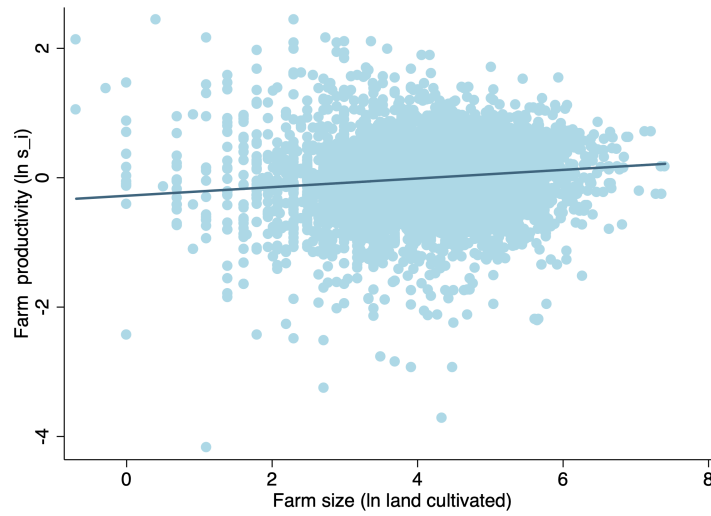


(b) Farm productivity ($\ln s_i$)

Figure B.3: Farm size and productivity - Bangladesh



(a) Land productivity ($\ln(Y/T)$)



(b) Farm productivity ($\ln s_i$)

Table B.1: Replication of Table 8: Correcting by DRS and market distortions countries

	Peru			Tanzania			Bangladesh		
	ln(output per ha.)			ln(output per ha.)			ln(output per ha.)		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
ln(land cultivated)	-0.533***	-0.533***	-0.286***	-0.403***	-0.403***	-0.152***	-0.103***	-0.103***	0.032***
$\beta + \gamma - 1$	(0.012)	(0.012)	(0.030)	(0.019)	(0.019)	(0.020)	(0.012)	(0.012)	(0.010)
ln(labor/land)			0.259***			0.447***			0.476***
$\gamma(1 - \alpha)$			(0.029)			(0.017)			(0.016)
Relax CRS assumption		Yes	Yes		Yes	Yes		Yes	Yes
Add input ratio			Yes			Yes			Yes
Assumed γ	1.000	0.384	0.384	1.000	0.691	0.691	1.000	0.904	0.904
Implied β	-0.533	0.083	0.330	-0.403	-0.094	0.157	-0.103	-0.014	0.128
No. obs.	11,359	11,359	11,357	7,899	7,899	7,890	6,506	6,506	6,506
R-squared	0.384	0.205	0.213	0.287	0.234	0.334	0.224	0.201	0.360

Notes: Robust standard errors in parentheses. Standard errors are clustered at the household level. * denotes significant at 10%, ** significant at 5% and *** significant at 1%. Results replicate columns 1-3 of Table 8. Regressions includes same controls as baseline results in Table 7. Assumed γ obtained from estimation of production function.

Table B.2: Robustness checks of yield-size relationship

	Peru			Tanzania			Bangladesh		
	ln(output per ha.)			ln(output per ha.)			ln(output per ha.)		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
ln(land cultivated)	-0.759*** (0.014)		-0.286*** (0.030)	-0.613*** (0.031)		-0.152*** (0.020)	-0.213*** (0.025)		0.040*** (0.010)
ln(land available)		-0.498*** (0.012)			-0.363*** (0.020)			-0.083*** (0.011)	
ln(labor/land)			0.259*** (0.029)			0.447*** (0.017)			0.476*** (0.016)
Household FE	Yes	No	No	Yes	No	No	Yes	No	No
No. obs.	11,359	11,359	11,357	7,899	7,899	7,890	6,506	6,506	6,506
R-squared	0.384	0.205	0.213	0.172	0.272	0.379	0.052	0.218	0.378

Notes: Robust standard errors in parentheses. Standard errors are clustered at the household level. * denotes significant at 10%, ** significant at 5% and *** significant at 1%. Results replicate columns 2-4 of Table 2. Regressions includes same controls as baseline results in Table 7. Column 1 also adds household fixed effects. Columns 3, 6 and 9 use the production function approach, while other columns use the yield approach.

Table B.3: Robustness checks of farm productivity-size relationship

	Peru			Tanzania			Bangladesh		
	ln(output per ha.)			ln(output per ha.)			ln(output per ha.)		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
ln(land cultivated)	0.183*** (0.011)	0.136*** (0.011)	0.176*** (0.011)	0.163*** (0.017)	0.201*** (0.017)	0.154*** (0.016)	0.078*** (0.010)	0.111*** (0.020)	0.083*** (0.009)
Prod. function used used to estimate s_i	CD by department	CD + IV	Translog	CD by region	CD + IV	Translog	CD by division	CD + IV	Translog
No. obs.	11,364	11,364	11,364	7,894	7,055	7,894	6,525	6,525	6,525
R-squared	0.301	0.333	0.314	0.868	0.450	0.576	0.430	0.246	0.234

Notes: Robust standard errors in parentheses. Standard errors are clustered at the household level. * denotes significant at 10%, ** significant at 5% and *** significant at 1%. Results replicate columns 4-6 of Table 3. Regressions includes same controls as baseline results in Table 7. No. of departments in Peru = 24. No. regions in Tanzania=26. No. divisions in Bangladesh=7.

Table B.4: Yields and labor productivity by farm size – United States

Farm size (acres)	Average farm size	Farm distribution (%)	Land share (%)	Value added per acre	Value added per worker
1–9	4.8	13.4	0.1	23.3	1.0
10–49	25.4	28.5	1.6	6.6	1.5
50–69	58.1	6.6	0.9	4.7	2.3
70–99	82.2	8.0	1.5	3.8	3.0
100–139	116.0	7.3	1.9	3.0	3.3
140–179	157.4	5.7	2.0	2.6	3.8
180–219	197.7	3.6	1.6	2.9	5.0
220–259	238.0	2.8	1.5	2.6	5.4
260–499	357.8	9.0	7.3	2.6	7.5
500–999	696.6	6.5	10.3	2.8	13.3
1,000–1,999	1376.6	4.3	13.4	2.4	19.3
2,000+	6103.4	4.2	57.7	1.0	22.7

Notes: Value added per acre and value added per worker are normalized relative to the lowest value. Data is from the 2017 US Census of Agriculture, Table 71, Summary by Size of Farm. Value added and adjusted farm labor are computed following Adamopoulos and Restuccia (2014).