The Farm Size-Productivity Relationship in Tanzania: Preliminary Findings

Ayala Wineman
Thomas S. Jayne

Presented at the Farm Size/Farm Productivity Conference, organized by Economic Research Service, USDA, Washington, DC, February 1-2, 2017

Hypotheses

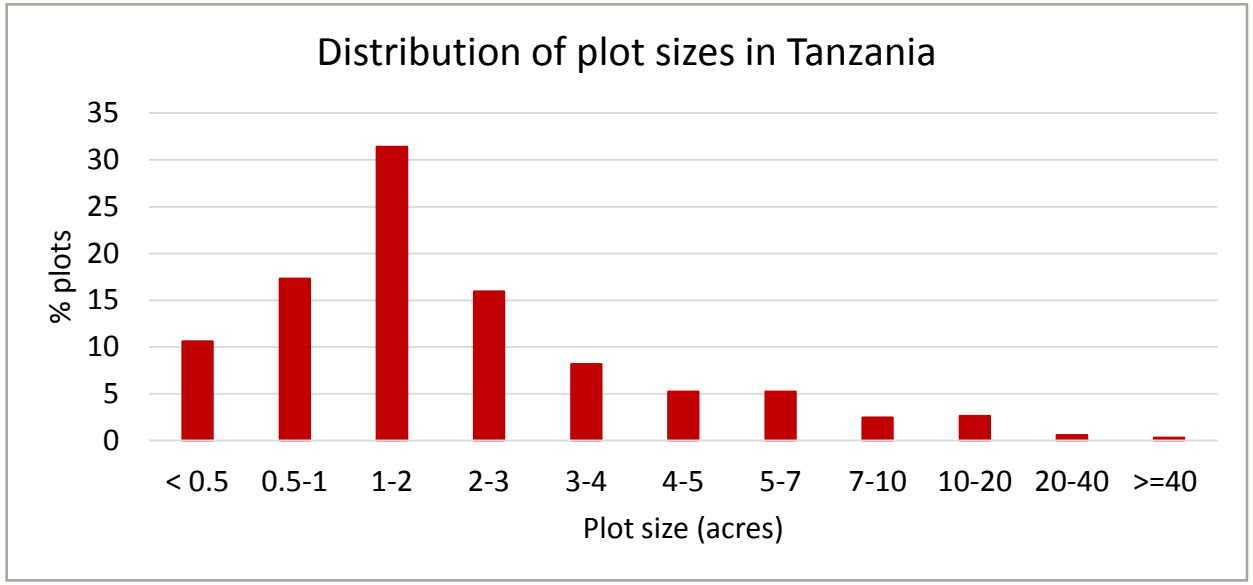
1. IR is a function of market failures.	It will disappear when we account for local levels of land or labor market activity.
2. IR is a function of plot-level characteristics.	It will disappear when we control for time-invariant plot fixed effects.
3. IR is a function of crop mix on small farms/ plots.	It will disappear when we account for crop mix in adequate detail.

Data

LSMS (NPS) Tanzania 2008/09, 2010/11, 2012/13

Number of plot-level observations

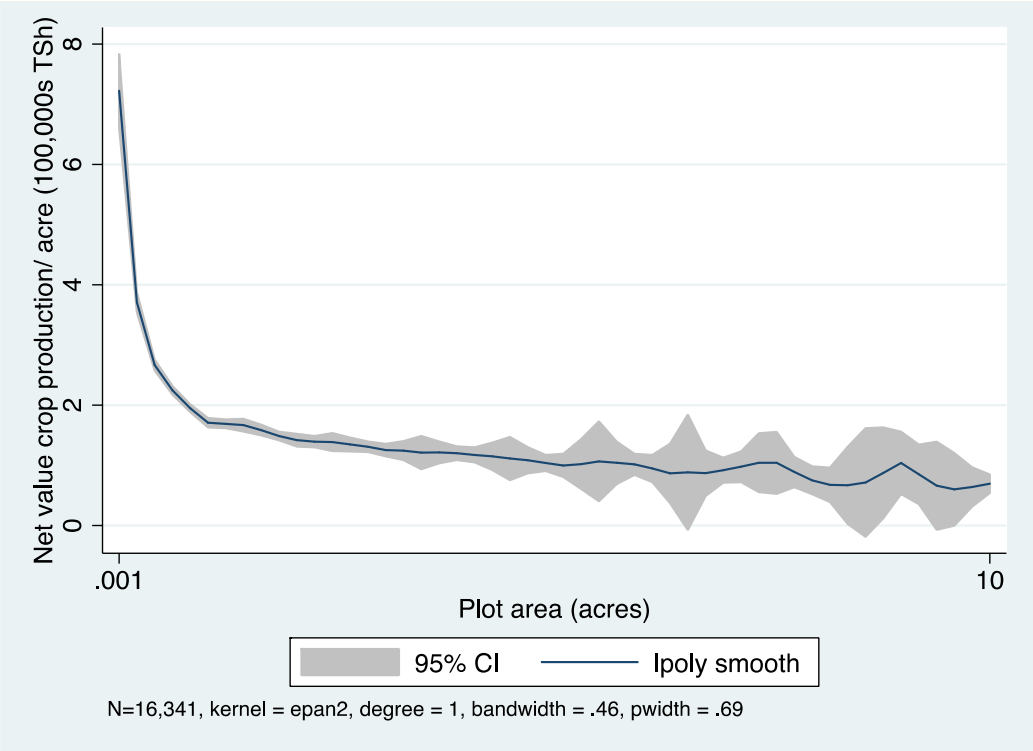
	Info on area and net value of crop production	Complete info for all RHS variables	Plots tracked from year 2009, present in all 3 survey waves with complete info in all waves
2008/09	4,734	4,401	2,370
2010/11	5,412	4,905	2,370
2012/13	6,635	6,187	2,370
Total	16,781	15,493	7,110
Sample restrictions		≤ 50 acres = 15,455	≤ 50 acres = 7,083



95th percentile:
7.20 acres (2.91 ha)

Relationship between plot area and crop revenue

Non-parametric polynomial regression



For visual clarity, sample excludes plots greater than 10 acres.

These coefficients represent the ***slope*** at this section of the plot-size spectrum.

Linear piecewise (spline) regression

Net value crop production/ acre	Coef	P-value
< 0.5 acres	-5.31***	0.00
0.5-1	-0.88***	0.00
1-1.5	0.15	0.42
1.5-2	-0.50***	0.009
2-3	-0.13*	0.09
3-4	-0.09	0.31
4-5	-0.17*	0.08
5-7	-0.06	0.28
7-10	-0.07*	0.06
10-20	-0.02**	0.02
20-40	0.0004	0.96
≥ 40 acres	-0.002**	0.03
Constant	4.83***	0.00
Observations	16,781	
R-squared	0.08	

*** p<0.01, ** p<0.05, * p<0.1

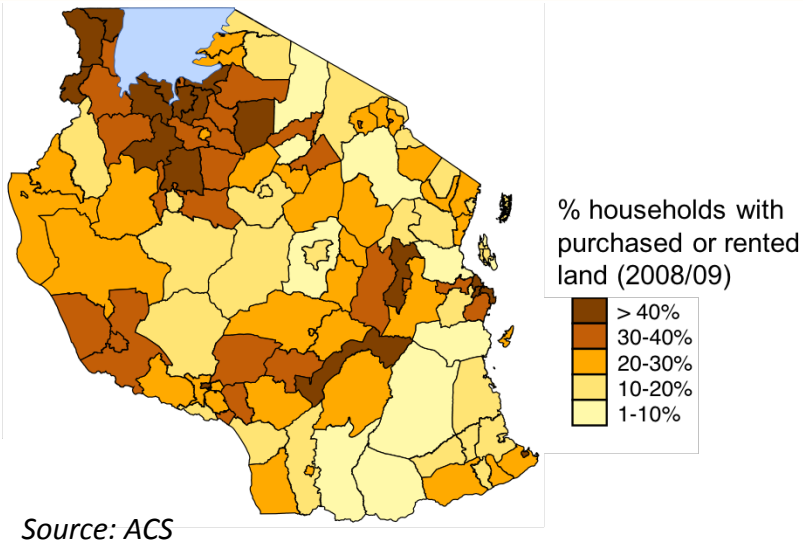
Relationship between plot area and crop revenue – Regression analysis (pooled OLS) –

	(1)	(2)	(3)	(4)	(5)
	Dependent variable: Net value crop production/ acre (100,000s TSh)				
Area (acres, estimated)	-0.14***	-0.29***	-0.12***	-0.05***	-0.04***
Area^2		0.01***			
1=Plot is right at residence			0.45***	0.33***	0.54***
Distance from plot to home (km)			-0.001***	-0.001***	-0.001
Distance from plot to road (km)			-0.03***	-0.03***	0.01
Distance from plot to market (km)			-0.01**	-0.002	-0.01
1= Problems with erosion on plot			-0.11*	-0.12**	0.09
1= Soil quality is 1 out of 3 (best)			0.37***	0.32***	0.34***
1= Soil quality is 3 out of 3 (worst)			-0.46***	-0.41***	-0.19
1= Slope is 'flat'			0.05	0.03	0.07
1= Slope is 'steep'			0.08	0.20*	0.19
Population density (persons/km^2)			0.000**	0.000***	
1= Plot cultivated in both seasons				0.46***	0.52***
1= Plot was irrigated (≥ 1 season)				1.67***	2.05***
Kgs manure/ acre				0.002***	0.001***
Kgs fertilizer/ acre				0.02***	0.01
Labor days/ acre (both seasons)				0.01***	0.01***
Region and Year Fixed Effects			Y	Y	
Household-Year Fixed Effects					Y
Constant	1.90***	2.13***	0.85***	0.32***	0.55***
Slope on area=0 at this value:		22.35 acres			
% Plots larger than this value:		0.61%			
Observations	15,455	15,455	15,455	15,455	12,801
R-squared	0.030	0.044	0.095	0.215	0.369

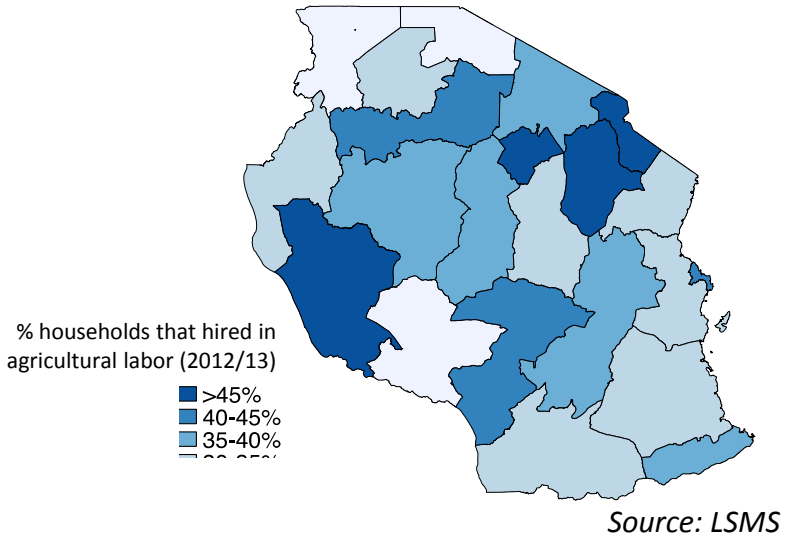
*** p<0.01, ** p<0.05, * p<0.1

Includes plots ≤ 50 acres

Relationship between plot area and crop revenue – controlling for local market activity –



Source: ACS

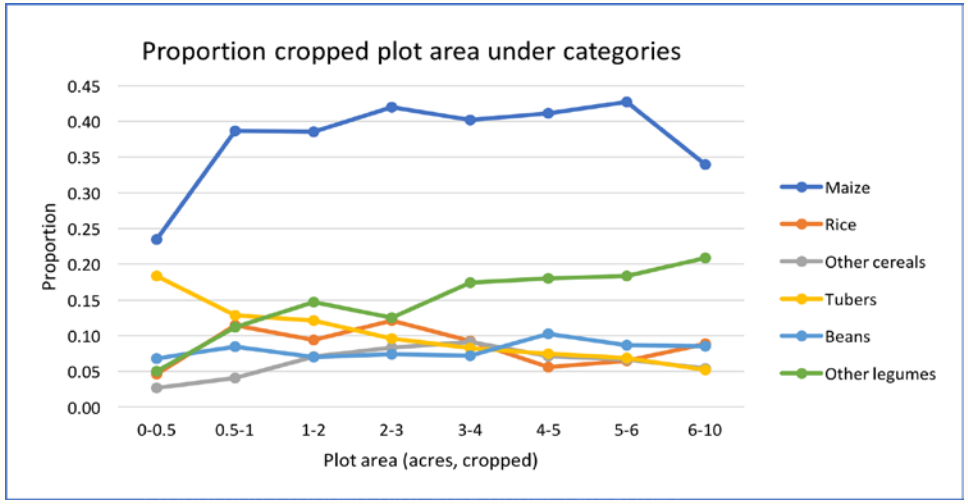


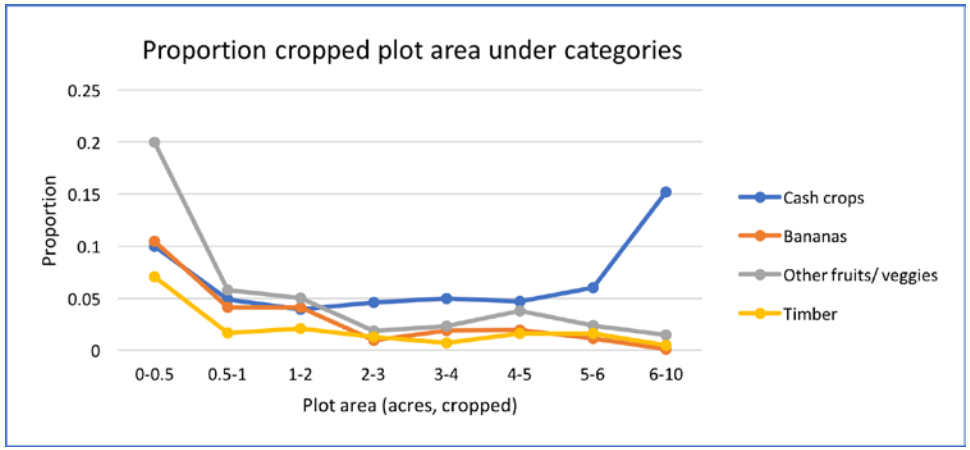
Source: LSMS

	(1)	(2)	(3)	(4)	(5)	(6)
	Dependent variable: Net value crop production/ acre (100,000s TSh)					
Area (acres, estimated)	-0.05***	-0.09***	-0.05***	-0.05***	-0.05***	-0.07***
Labor market activity (proportion)	-0.37	-0.63				
Area * Labor market activity		0.12**				
Land market activity			0.51*	0.49		
Area * Land market activity				0.01		
Land *rental* market activity					-0.02	-0.44
Area * Rental market activity						0.19***
Other control variables	Y	Y	Y	Y	Y	Y
Region and year fixed effects	Y	Y	Y	Y	Y	Y
market activity level at which δy/δa=0:		0.75		>1		0.35
% Households beyond this point		0.09%		0%		24.39%

N = 15,455

Plot size and crop revenue – controlling for crop mix –





Pooled OLS

Dependent variable: Net value crop production/ acre (100,000s TSh)	(1) Regressors = Proportion of cropped *area* (main season)	(2) Regressors = Proportion of *value* of crop production (both seasons)
Area (acres, estimated)	-0.05***	-0.06***
Rice	1.42***	1.84***
Other cereals (maize = omitted)	0.02	0.21***
Tubers	0.48***	0.96***
Beans	0.29***	0.24**
Other legumes	0.55***	0.80***
Cash crops	1.39***	1.99***
Bananas	2.33***	2.71***
Other fruits and vegetables	1.12***	1.40***
Spices	5.06***	5.58***
Sugarcane	3.22***	3.57***
Timber	0.10	2.69***
Other control variables	Y	Y
Region and Year Fixed Effects	Y	Y
Observations	15,455	15,455
R-squared	0.27	0.30

Relationship between plot area and crop revenue – Correlated Random Effects (CRE) regressions –

$$Y_{ph} = \alpha + \beta A_{ph} + \boldsymbol{X_{ph}\delta} + \boldsymbol{T_t\gamma} + \theta_p + \varepsilon_{ph}$$

- Y_{ph}: Net value crop production/ acre
- A_{ph}: Plot area (time-invariant)
- $\boldsymbol{X_{ph}\delta}$: Plot characteristics, Plot management, Region fixed effects, Year fixed effects
- $\boldsymbol{T_t\gamma}$: Plot fixed effects (mean values of all time-varying regressors)

	(1)	(2)	(3)	(4)
	Net value crop production/ acre (100,000s TSh)			
Area (acres, time-invariant value estimated)	-0.02***	-0.07*	-0.03***	-0.03***
Labor market activity (proportion)		-0.17		
Area * Labor market activity		-0.08		
Land rental market activity (time-invariant)			0.20	
Area * Rental market activity (time-invariant)			0.04	
Crop mix variables (proportions of crop value)				Y
Other control variables	Y	Y	Y	Y
Region and year fixed effects	Y	Y	Y	Y
Mean values of all time-varying regressors	Y	Y	Y	Y
Observations	7,083	7,083	7,083	7,083
Adjusted R-squared	0.25	0.25	0.25	0.34

*** p<0.01, ** p<0.05, * p<0.1; Inverse probability weights to address likelihood of plot remaining in sample

Preliminary Conclusions

- IR is **persistent** in plot-level analysis.
- IR evident along spectrum of plot sizes.
- Crop mix, unobserved plot effects do not seem to (fully) explain the IR
- Pooled OLS indicates IR intensity is at least partially correlated with local levels of labor and land market activity. However, this relationship does not seem to persist in a CRE analysis.

What's next?

- Treatment of plot measurement error
- Further consideration of heterogeneity along plot size spectrum (e.g., market activity interactions)

Pooled OLS from slide #2 (with all plots including those > 50 acres)

	(1)	(2)	(3)	(4)	(5)
	Dependent variable: Net value crop production/ acre (100,000s TSh)				
Area (acres, estimated)	-0.05***	-0.10***	-0.04***	-0.02***	-0.009
Area^2		0.0004***			
Control variables			Y	Y	Y
Region and Year Fixed Effects			Y	Y	
Household-Year Fixed Effects					Y
Slope on area=0 at this value:		124.74 acres			
% Plots larger than this value:		0.03%			
Observations	15,493	15,493	15,493	15,493	12,829

*** p<0.01, ** p<0.05, * p<0.1

Pooled OLS from slide #2 (excluding plots < 0.5 acres)

	(1)	(2)	(3)	(4)	(5)
	Dependent variable: Net value crop production/ acre (100,000s TSh)				
Area (acres, estimated)	-0.04***	-0.08***	-0.03***	-0.02***	-0.01
Area^2		0.0003***			
Control variables			Y	Y	Y
Region and Year Fixed Effects			Y	Y	
Household-Year Fixed Effects					Y
Observations	13,635	13,635	13,635	13,635	11,214

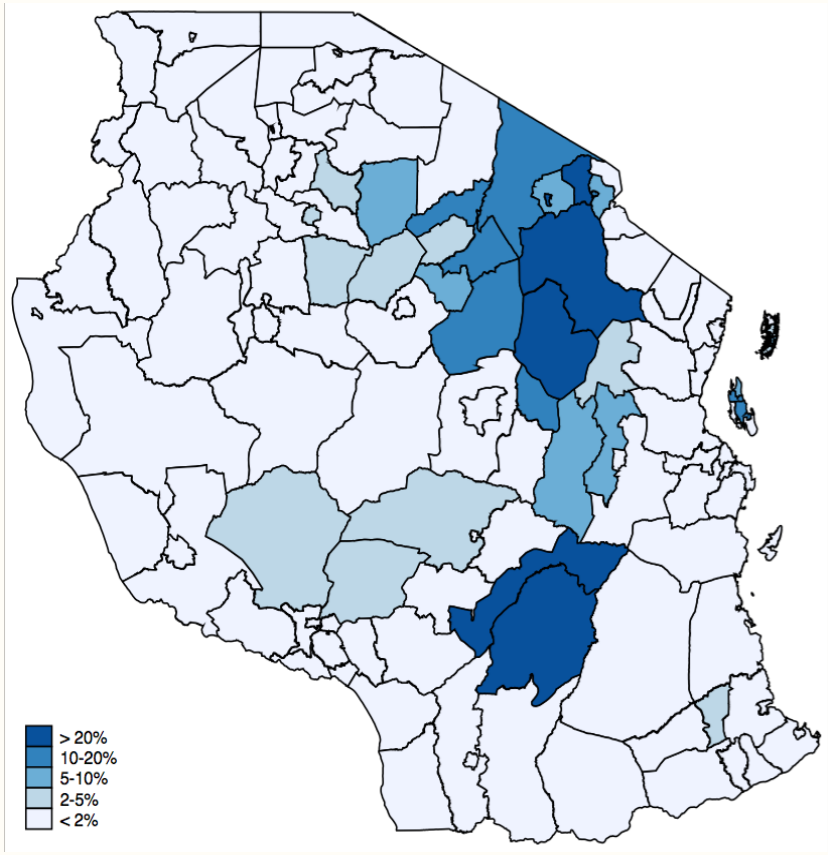
*** p<0.01, ** p<0.05, * p<0.1

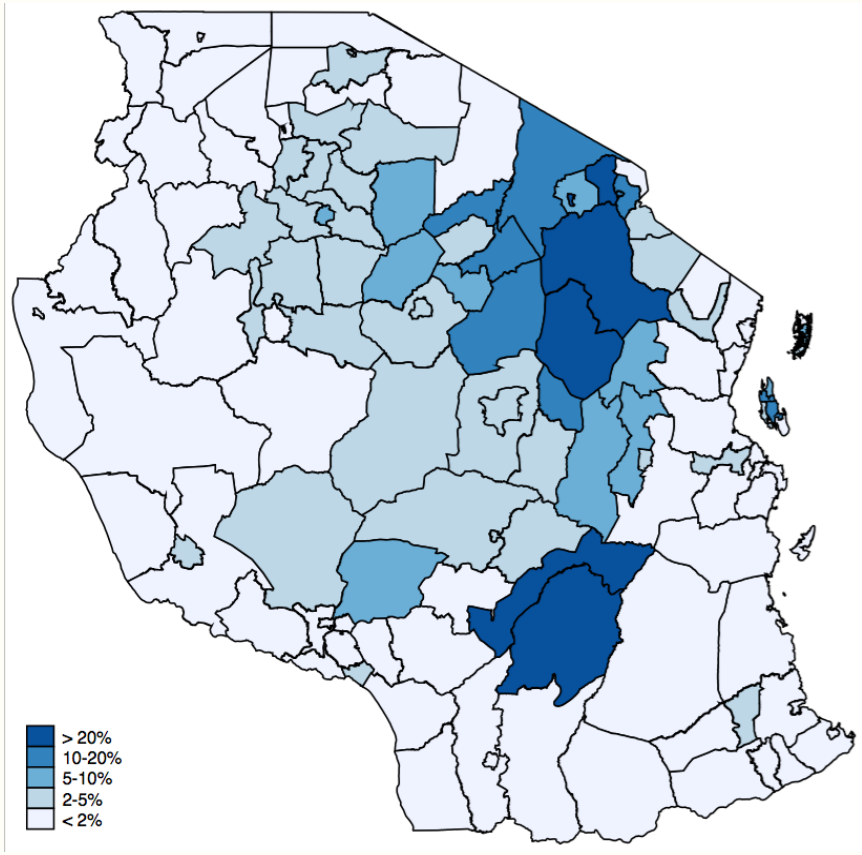
From slide #4 (excluding plots < 0.5 acres)

	(1)	(2)	(3)	(4)	(5)	(6)
	Dependent variable: Net value crop production/ acre (100,000s TSh)					
Area (acres, estimated)	-0.02***	-0.01	-0.02***	-0.01	-0.02***	-0.02**
Labor market activity (proportion)	0.003	-0.05				
Area * Labor market activity		-0.02				
Land market activity			1.00***	1.04***		
Area * Land market activity				-0.01		
Land *rental* market activity					0.90**	0.85**
Area * Rental market activity						0.02
Other control variables	Y	Y	Y	Y	Y	Y
Region and year fixed effects	Y	Y	Y	Y	Y	Y
Slope on area=0 at market activity level:		N/A		N/A		0.55
% Households beyond this point						2.78%

Thank you

Percent of cropping households that rented/ borrowed a tractor (left) or used a tractor (right)

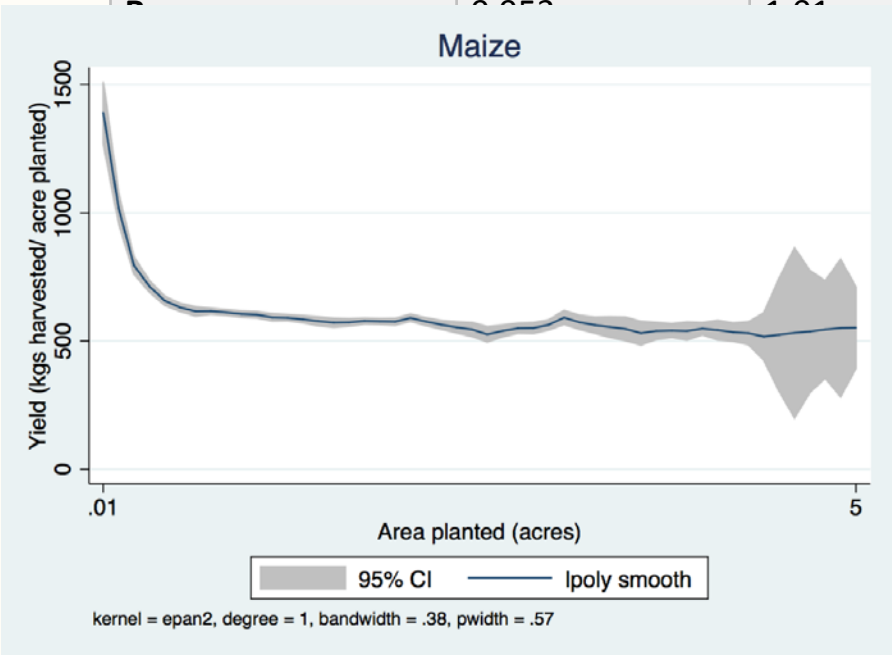


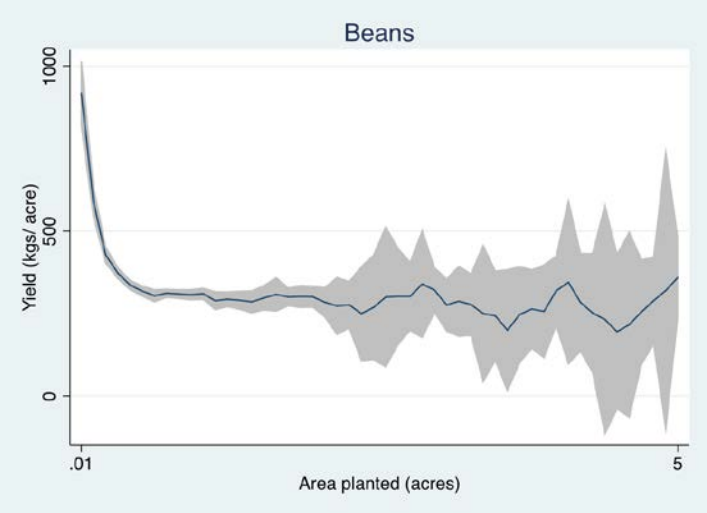


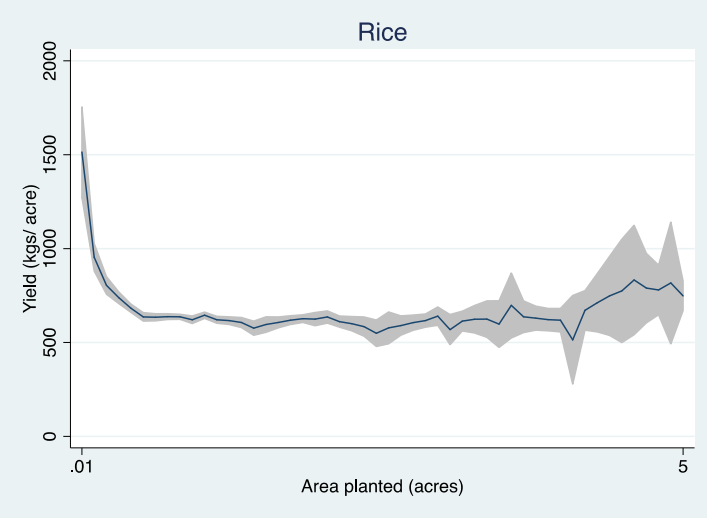
Relationship between area planted and crop yield (specific crops, Agricultural Sample Census 2008/09)

Non-parametric polynomial regressions

	Observations	Area plant Mean
Maize	29,935	2.27
Rice	9,139	1.90

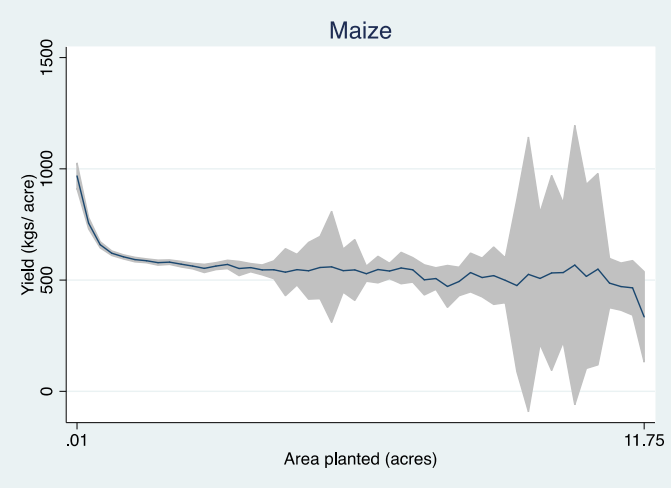


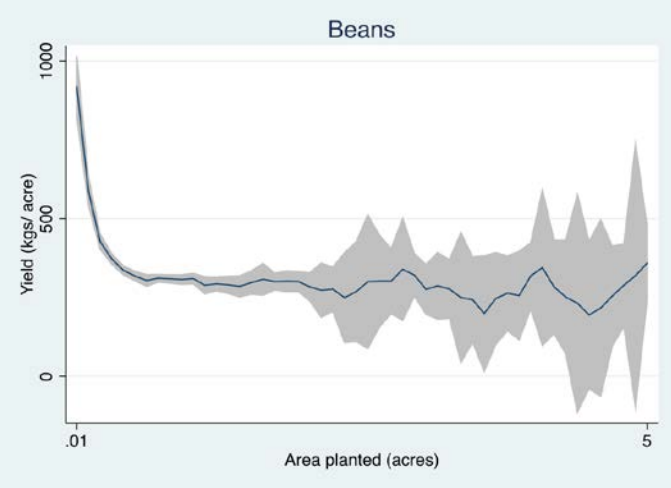


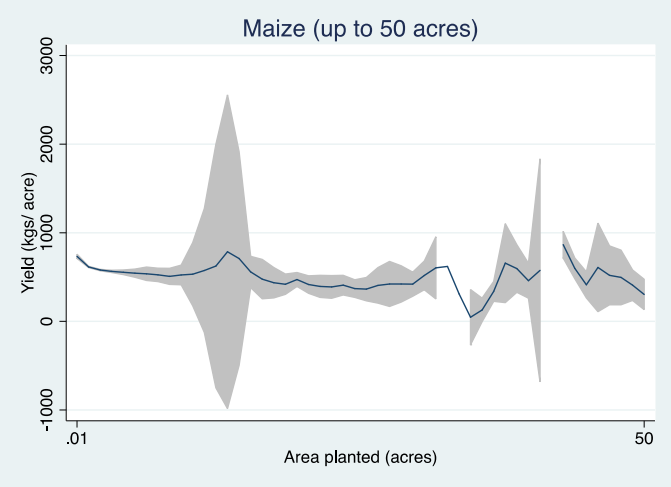


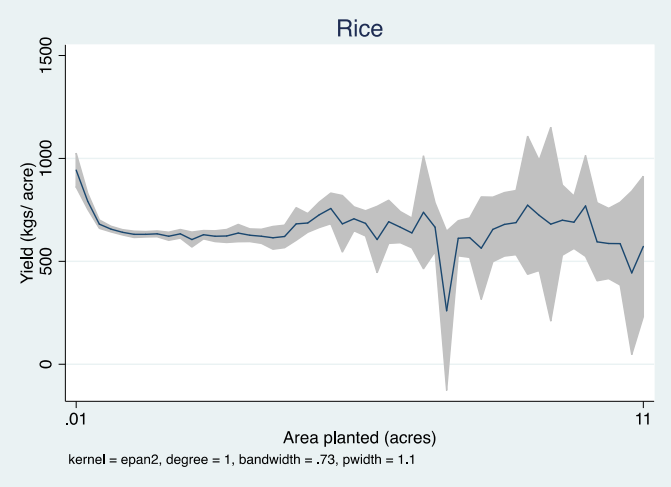
Relationship between area planted and crop yield (specific crops, Agricultural Sample Census 2008/09)

Non-parametric polynomial regressions, x-axis extends up to 99th percentile



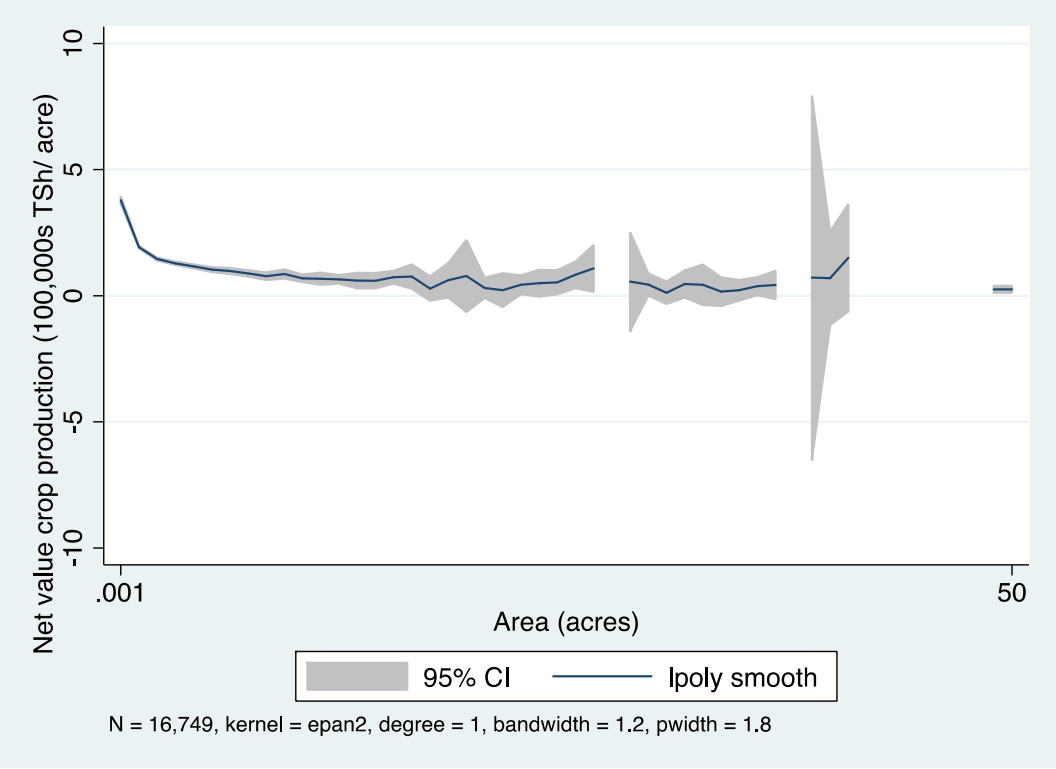






Sensitivity Analyses
with different sample restrictions (LSMS data)

Non-parametric polynomial regression (including plots ≤ 50 acres)



Note: We use the rule-of-thumb bandwidth selector in Stata