EVANS SCHOOL OF PUBLIC POLICY AND GOVERNANCE
UNIVERSITY of WASHINGTON

How Common Crop Yield Measures Misrepresent Productivity Among Smallholder Farmers

C. Leigh Anderson, Travis Reynolds, and Elysia Slakie*

August 10th, 2015
29th ICAE, Università degli Studi di Milano

* The authors thank the Bill & Melinda Gates Foundation for supporting this work

Contents

EVANS SCHOOL OF PUBLIC POLICY AND GOVERNANCE
UNIVERSITY of WASHINGTON

Increasing Agricultural Productivity

WHAT WE INTEND

Increasing Productivity

Increasing farm productivity regarded as prerequisite for improvement of rural livelihoods and development in low-income countries, particularly for SSA (Pingali, 2011)

- FAO: "improve agricultural productivity"
- World Bank: "increasing agricultural productivity" (75% of ag lending)
- USAID: "increased productivity" key to "inclusive agriculture-led growth"
- BMGF Foundation: "increasing agricultural productivity in a sustainable way"

Governments, non-profits, and others have invested billions in pursuit of higher productivity for smallholder farmers.

EVANS SCHOOL OF PUBLIC POLICY AND GOVERNANCE
UNIVERSITY of WASHINGTON

Crop Yield as Proxy for Smallholder Productivity

WHAT WE MEASURE

Productivity Literature

- **Agricultural Productivity Measures**
  - defined in several ways in the literature, including output per unit of input (total factor productivity), farm yield by crop or total output per hectare, and output per worker (total or partial factory productivity; Fuglie, 2008; Alston et al., 2010)
  - meta-studies and literature reviews (Schneider and Gugerty, 2011, Irz et al., 2001; Mellor, 1999; Thirtle et al., 2001); macrolevel studies including Ravallion and Datt (1996, 1998) and Timmer (1995, 1997). Micro-level evidence includes several cross-country studies (Byerlee et al., 2009; Minten & Barrett, 2008; Muyanga et al., 2010)

- **Data and Measurement Issues**
  - Administratively reported production estimates, such as those compiled and reported by the FAO, may be fraught with statistical and political error (Sandefur & Glassman, 2015; Jerven, 2014)
  - National-level yield estimates may differ starkly from yields realized by any given smallholder farmer, regionally, or within a sub-population (Craig et al., 1997)
  - Missing markets and missing data: prices, wages, natural resource use etc.

Common Crop Yield

Common crop yield is widely used to proxy for smallholder farm productivity.

$$\text{Common crop yield} = \frac{\sum \text{Quantity harvested in kg}}{\sum \text{Area harvested in ha}}$$

- Similar biases with administrative and/or household (survey) level data
- Similarly national average masks regional or household-level variation
- Additional measurement error with hh survey data self-reporting bias (Carletto et. al, 2013a and b; De Groote and Traorè)

Validity Issues

1. Using yield to proxy productivity

Common crop yield captures a single output from a single input at a single moment

- Use of common crop yield as the sole indicator ignores the value of multiple outputs and the costs associated with other inputs to farm production including labor, tools and environmental services (Reynolds et al., 2015; Cassidy et al., 2013; Alston et al., 2010; Ehui & Pender, 2005)

Quantity harvested: complicated by multi/inter-cropping and ongoing harvesting of crops such as cassava

Area harvested: common yield measurement is complicated by land factors such as irregular plot shapes and non-planted areas due to trees, stumps, anthills/termite mounds and other obstructions (Fermont & Benson, 2011; Casley & Kumar, 1988).

2. Using yield (land productivity) based on area harvested:

Plot area harvested may be substantially smaller than plot area planted due to poor germination, damage from pests or disease, floods, labor constraints, or lack of market opportunities – all common circumstances for small scale farmers (Fermont & Benson, 2011).

Our empirical focus:

a. Do estimates of yield vary?

b. Do these differences matter (in directing resources)?

Are data and measurement errors random?

c. How do they matter (in which direction does the bias run)?

Does calculating yield over area harvested overestimate mean crop yield, instead measuring "productivity among the most productive."

EVANS SCHOOL OF PUBLIC POLICY AND GOVERNANCE
UNIVERSITY of WASHINGTON

FIndings: Rice in Tanzania

DOES IT DIFFER & DOES IT MATTER?

EVANS SCHOOL OF PUBLIC POLICY AND GOVERNANCE
UNIVERSITY of WASHINGTON

Rice Yield Estimates 2001-2011

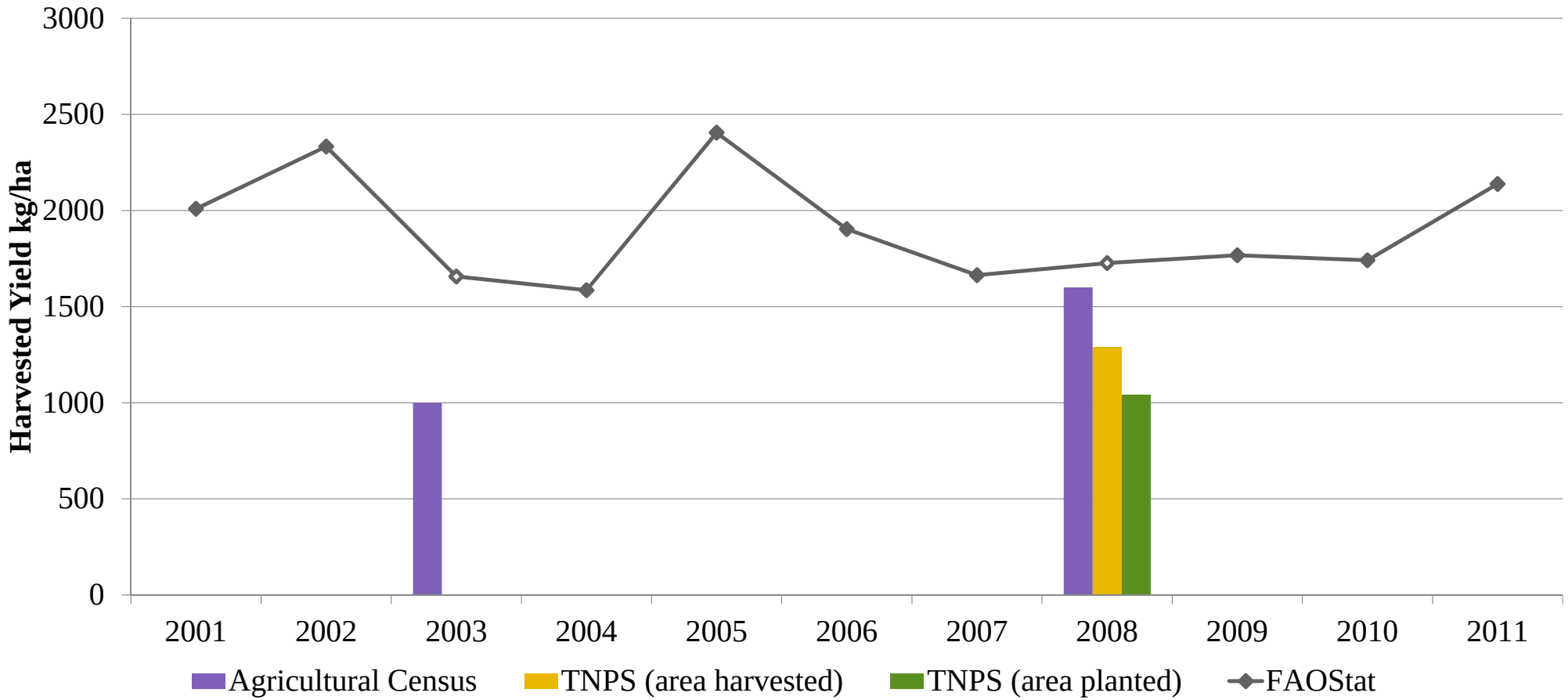
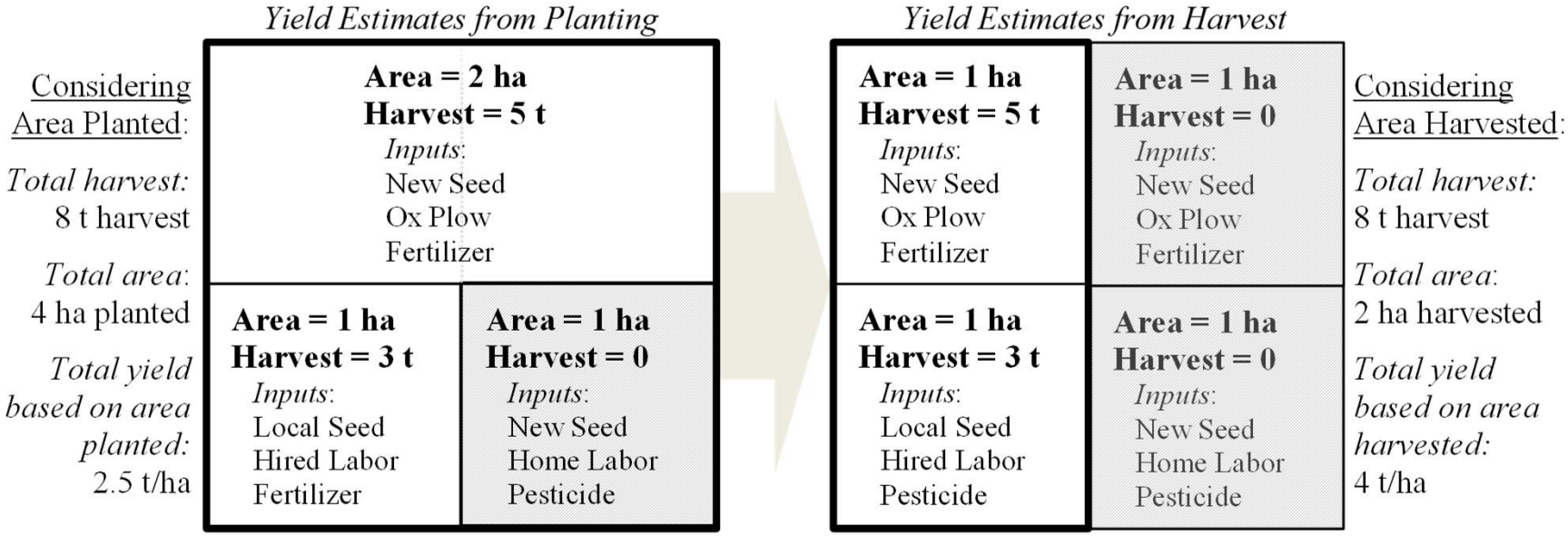


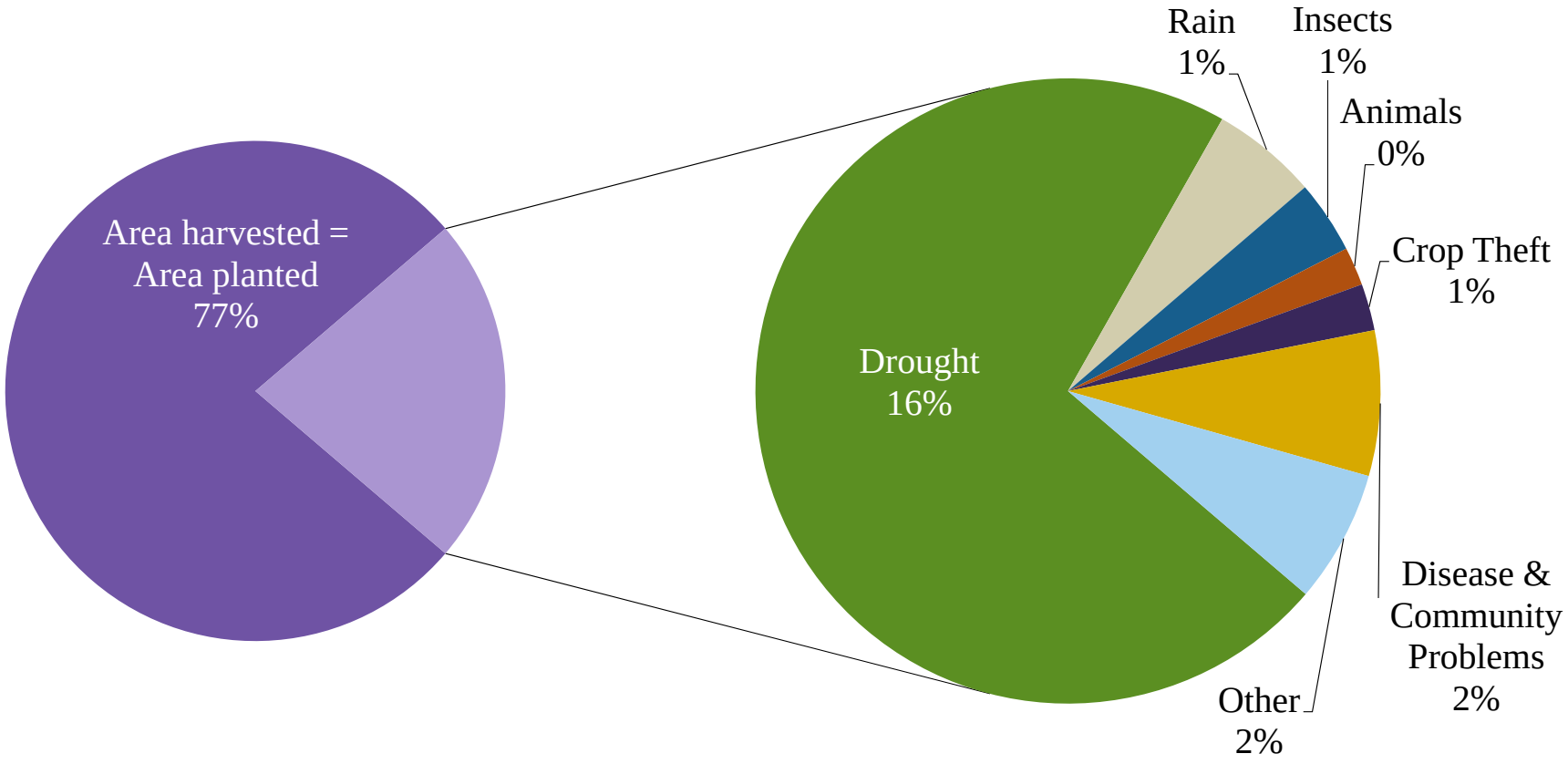
Figure 1. How Common Yield Measures Can Misrepresent Crop Yield and Bias Marginal Yield Gain Estimates. (Hypothetical example: A farmer plants 4 hectares of crop, 2 hectares fail and are not harvested.)



Bias in Marginal Yield Gain Estimates: *In this hypothetical example, when considering production per area planted (on left) the best combination is local seed/hird labor/fertilizer, which offers a mean yield of 3 t/ha[planted] which is higher than 2.5 t/ha[planted] from the new seed/ox plow/fertilizer combination. However, when considering production per area harvested (common yield, on right) the best combination seems to be new seed/ox plow/ fertilizer, with 5 t/ha [harvested]. Ignoring the failed crop plots results in biased estimates of management-based yield gains.*

EVANS SCHOOL OF PUBLIC POLICY AND GOVERNANCE
UNIVERSITY of WASHINGTON

Farmer-Reported Reasons for Loss in Area between Planting and Harvesting



OLS Regression Results

Table 4. OLS regression results for yield by area harvested versus yield by area planted

		Model 1: ***Yield by area harvested***		**Model 2:** ***Yield by area planted***	
Constraint type	Description	N: 376 R^2: 0.4121		N: 376 R^2: 0.4973	
		Coefficient	**p-value**	**Coefficient**	**p-value**
Abiotic	No or slight constraints to soil nutrient availability	**726.45*** **	**0.006**	**315.11***	**0.069**
	No or slight constraints to soil workability=1	204.36	0.244	120.76	0.398
	Annual mean temperature 1960-1990 (10 degrees C)	**-17.46*** **	**0.009**	**-10.88***	**0.061**
	Rainfall more than 50 mm higher than 9 year average=1	-185.49	0.499	**-651.80** **	**0.034**
	Rainfall less than 50 mm higher than 9 year average=1	**-371.68** **	**0.029**	**-297.95***	**0.064**
Biotic	Improved variety seed=1	-214.38	0.242	-32.40	0.844
	Farmer reported losses due to birds=1	36.63	0.814	6.86	0.959
	Farmer reported losses due to other causes=1	53.72	0.750	-117.30	0.367
	Pesticide, herbicide, or fungicide use on plot=1	452.68	0.160	279.54	0.157
Management	Inorganic fertilizer use on plot=1	305.66	0.188	347.78	0.117
	Rice intercropped on plot=1	-220.97	0.182	-176.39	0.232
	Number of years the plot was left fallow	**214.37***	**0.056**	94.82	0.121
	Plot size in hectares	**-317.11** **	**0.022**	**-271.03*** **	**0.005**
	Number of plots owned by the household	-45.90	0.573	-88.34	0.170
Socioeconomic	Zanzibar=1	32.57	0.889	74.99	0.730
	Female head of household=1	-102.04	0.630	**-211.85** **	**0.042**
	Age of head of household	-8.20	0.168	-4.64	0.300
	Years of education of head of household	**56.34** **	**0.031**	**31.22** **	**0.049**
	Household labor days per hectare	**1.16*** **	**0.000**	**0.92*** **	**0.000**
	Hired labor days per hectare	**8.62*** **	**0.000**	**9.08*** **	**0.000**
	Household owned/rented ox, ox plough, ox planter, or ox cart=1	**1,195.55*** **	**0.000**	**642.72*** **	**0.000**
	Household received advice from any source=1	-224.41	0.181	**-250.60***	**0.061**
	Household sold rice=1	187.79	0.522	**571.53*** **	**0.000**
	Constant	**767.20***	**0.106**	**896.56** **	**0.021**

*** p<0.01, ** p<0.05, * p<0.1 | Values in bold represent significant variables. Shading indicates significance in one model but not in the other.

Conclusions

1. The value of hh survey data

Subpopulations: national and average yield estimates mask considerable variation, as do farm and per hh member estimates

- Smallholder yields differ drastically from national trends
- Yield per hh member in female headed hhs can be equivalent to male headed hhs.

Remediation: Measurement error is likely easier to fix than political error.

2. The potentially real consequences of measurement error

Yield v. productivity: missing multi-cropped outputs and inputs other than land

Common yield: measuring yield by area harvested produces very different estimates than by area planted

- Analyses based on these results will lead to different conclusions regarding factors that explain yield variability
- Cases where bias is non-random across: differences in farming practices, agro-ecological zones, gender and levels of poverty may run counter to intent.

EVANS SCHOOL OF PUBLIC POLICY AND GOVERNANCE
UNIVERSITY of WASHINGTON

Evans School Policy Analysis & Research Group (EPAR)

Professor Leigh Anderson, Principal Investigator

Jessica Rudder, Mathew Lane, Katie Panhorst Harris, Elysia Slakie, Travis Reynolds, & C. Leigh Anderson

EPAR's innovative student-faculty team model is the first University of Washington partnership to provide rigorous, applied research and analysis to the Bill and Melinda Gates Foundation. Established in 2008, the EPAR model has since been emulated by other UW Schools and programs to further support the foundation and enhance student learning.

Please direct comments or questions about this research to Leigh Anderson at eparx@u.washington.edu.

EVANS SCHOOL OF PUBLIC POLICY AND GOVERNANCE
UNIVERSITY of WASHINGTON

Regression Results: Rice and Maize, 2010 and 2012

Household-level data analysis from the Tanzania Living Standards Measurement Study – Integrated Surveys on Agriculture

EVANS SCHOOL OF PUBLIC POLICY AND GOVERNANCE
UNIVERSITY of WASHINGTON

Rice, Tanzania, 2010

		Model 1: *Yield by area harvested*		Model 2: *Yield by area planted*	
Constraint type	Description	N: 625 R^2: 0.32		N: 625 R^2: 0.34	
		Coefficient	**t statistic**	**Coefficient**	**t statistic**
Abiotic	No or slight constraints to soil nutrient availability=1	270.890	1.38	360.845	1.37
	No or slight constraints to soil workability=1	-206.793	1.09	-348.363	1.28
	Annual mean temperature 1960-1990 (10 degrees C)	-7.552	1.12	-3.840	0.58
	Rainfall more than 50 mm higher than 9 year average=1	-94.042	0.56	-94.588	0.52
	Rainfall less than 50 mm higher than 9 year average=1	-131.142	0.70	-102.167	0.44
Biotic	Improved variety seed=1	-379.191	1.72	-234.961	0.72
	Farmer reported losses due to birds=1	-41.011	0.29	-103.825	0.58
	Farmer reported losses due to other causes=1	-309.783	1.73	**-385.016***	**2.13**
	Pesticide, herbicide, or fungicide use on plot=1	82.545	0.45	82.157	0.27
Management	Inorganic fertilizer use on plot=1	193.983	0.72	380.399	1.34
	Rice intercropped on plot=1	0.825	0.00	-24.767	0.09
	Number of years the plot was left fallow	-27.089	0.31	67.021	0.59
	Plot size in hectares	-41.346	1.05	-50.009	1.51
	Number of plots owned by the household	38.685	0.73	6.316	0.10
Socioeconomic	Zanzibar=1	-175.091	0.73	-598.122	1.60
	Female head of household=1	-163.691	0.94	-19.694	0.11
	Age of head of household	-4.568	1.20	-1.694	0.40
	Years of education of head of household	18.488	0.93	25.716	1.26
	Household labor days per hectare	**0.586***	**2.45**	**2.647***	**2.49**
	Hired labor days per hectare	**8.014****	**6.64**	**8.314****	**8.16**
	Household owned/rented ox, ox plough, ox planter, or ox cart=1	-193.891	0.94	-77.900	0.31
	Household received advice from any source=1	126.496	0.60	83.755	0.36
	Household sold rice=1	**801.720****	**4.65**	**983.149****	**4.69**
	Constant	2865.34	1.87	1392.264	0.89

*** p<0.01, ** p<0.05, * p<0.1 | Values in bold represent significant variables. Shading indicates significance in one model but not in the other.

EVANS SCHOOL OF PUBLIC POLICY AND GOVERNANCE
UNIVERSITY of WASHINGTON

Rice, Tanzania, 2012

		Model 1: *Yield by area harvested*		Model 2: *Yield by area planted*	
Constraint type	Description	N: 749 R^2: 0.3181		N: 749 R^2: 0.3830	
		Coefficient	**t statistic**	**Coefficient**	**t statistic**
Abiotic	No or slight constraints to soil nutrient availability=1	-140.6	0.95	256.3	1.29
	No or slight constraints to soil workability=1	-131.3	0.62	-73.45	0.38
	Annual mean temperature 1960-1990 (10 degrees C)	-1.472	0.20	-9.643	1.30
	Rainfall more than 50 mm higher than 9 year average=1	348.1	1.92	-118.2	0.45
	Rainfall less than 50 mm higher than 9 year average=1	119.5	0.48	-353.4	1.26
Biotic	Improved variety seed=1	51.29	0.27	234.2	1.36
	Farmer reported losses due to birds=1	129.0	0.91	108.5	0.55
	Farmer reported losses due to other causes=1	-340.1	1.71	-223.0	1.05
	Pesticide, herbicide, or fungicide use on plot=1	**782.2****	**2.82**	**508.9***	**2.03**
Management	Inorganic fertilizer use on plot=1	432.0	1.34	214.6	0.7
	Rice intercropped on plot=1	**-538.2****	**3.27**	-13.05	0.03
	Number of years the plot was left fallow	**-337.9****	**2.63**	**-358.5****	**3..31**
	Plot size in hectares	-10.48	0.41	**-61.94****	**2.86**
	Number of plots owned by the household	**-137.4****	**2.89**	**-251.6*****	**4.08**
Socioeconomic	Zanzibar=1	**-933.6*****	**4.36**	**-846.7*****	**3.71**
	Female head of household=1	**-334.9***	**2.02**	-147.0	0.76
	Age of head of household	-4.944	1.32	-0.604	0.11
	Years of education of head of household	-12.95	0.63	8.644	0.39
	Household labor days per hectare	**1.077****	**2.87**	**3.232****	**3.2**
	Hired labor days per hectare	**6.706****	**3.09**	**8.041*****	**4.29**
	Household owned/rented ox, ox plough, ox planter, or ox cart=1	-22.49	0.13	398.8	1.57
	Household received advice from any source=1	379.4	1.31	319.7	1.19
	Household sold rice=1	**780.5*****	**5.66**	**1073.9*****	**6.46**
	Constant	1768.1	1.0	2998.1	1.72

*** p<0.01, ** p<0.05, * p<0.1 | Values in bold represent significant variables. Shading indicates significance in one model but not in the other.

EVANS SCHOOL OF PUBLIC POLICY AND GOVERNANCE
UNIVERSITY of WASHINGTON

Maize, Tanzania, 2010

		Model 1: *Yield by area harvested*		Model 2: *Yield by area planted*	
Constraint type	Description	N: 2251 R^2: 0.06		N: 2252 R^2: 0.21	
		Coefficient	**t statistic**	**Coefficient**	**t statistic**
Abiotic	No or slight constraints to soil nutrient availability=1	92.366	0.71	**270.521***	**2.20**
	No or slight constraints to soil workability=1	-128.348	1.67	-3.164	0.03
	Annual mean temperature 1960-1990 (10 degrees C)	-3.345	1.28	-1.133	0.53
	Rainfall more than 50 mm higher than 9 year average=1	193.260	1.58	112.120	1.07
	Rainfall less than 50 mm higher than 9 year average=1	63.635	0.73	129.464	1.41
Biotic	Improved variety seed=1	446.970	1.94	161.951	1.84
	Farmer reported losses due to birds=1	**-346.027****	**4.13**	**-333.848****	**3.28**
	Farmer reported losses due to other causes=1	198.673	1.67	54.743	0.67
	Pesticide, herbicide, or fungicide use on plot=1	-19.757	0.16	229.932	1.22
Management	Inorganic fertilizer use on plot=1	**256.100***	**2.48**	**635.290****	**3.99**
	Maize intercropped on plot=1	-163.167	1.55	-91.244	1.29
	Number of years the plot was left fallow	17.796	0.41	88.290	1.56
	Plot size in hectares	32.280	0.96	-18.330	1.25
	Number of plots owned by the household	-17.220	0.68	-13.218	0.66
Socioeconomic	Zanzibar=1	-379.277	1.80	**-1417.119****	**3.12**
	Female head of household=1	-77.494	0.88	-110.805	1.49
	Age of head of household	-1.748	0.51	**-4.241***	**2.28**
	Years of education of head of household	4.048	0.18	8.144	1.15
	Household labor days per hectare	0.268	1.49	**3.143****	**2.88**
	Hired labor days per hectare	1.676	1.29	**5.315****	**2.80**
	Household owned/rented ox, ox plough, ox planter, or ox cart=1	317.007	1.82	296.470	1.90
	Household received advice from any source=1	137.664	0.93	55.208	0.46
	Household sold maize=1	**412.045****	**5.74**	**418.129****	**5.85**
	Constant	**1464.814***	**2.00**	460.797	0.80

*** p<0.01, ** p<0.05, * p<0.1 | Values in bold represent significant variables. Shading indicates significance in one model but not in the other.

EVANS SCHOOL OF PUBLIC POLICY AND GOVERNANCE
UNIVERSITY of WASHINGTON

Maize, Tanzania, 2012

		Model 1: *Yield by area harvested*		Model 2: *Yield by area planted*	
Constraint type	Description	N: 2891 R^2: 0.0053		N: 2891 R^2: 0.1873	
		Coefficient	**t statistic**	**Coefficient**	**t statistic**
Abiotic	No or slight constraints to soil nutrient availability=1	874.5	1.23	113.5	1.40
	No or slight constraints to soil workability=1	-277.1	0.60	67.5	0.79
	Annual mean temperature 1960-1990 (10 degrees C)	0.279	0.06	-1.164	0.58
	Rainfall more than 50 mm higher than 9 year average=1	-1136.8	0.95	-94.6	0.92
	Rainfall less than 50 mm higher than 9 year average=1	-792.1	0.95	-187.6	1.77
Biotic	Improved variety seed=1	-627.7	0.88	3.223	0.04
	Farmer reported losses due to birds=1	-333.0	1.00	-145.2	1.00
	Farmer reported losses due to other causes=1	-1201.4	1.56	**-170.5***	**2.22**
	Pesticide, herbicide, or fungicide use on plot=1	-168.5	0.59	108.2	0.47
Management	Inorganic fertilizer use on plot=1	244.1	1.32	**505.2*****	**4.62**
	Maize intercropped on plot=1	-1063.0	1.19	**-274.9****	**2.94**
	Number of years the plot was left fallow	-224.4	1.37	-91.18	1.95
	Plot size in hectares	-68.52	1.31	-1.675	0.16
	Number of plots owned by the household	-295.2	1.13	-26.76	0.68
Socioeconomic	Zanzibar=1	-887.9	0.82	-585.0	1.70
	Female head of household=1	-981.7	1.09	-68.18	0.84
	Age of head of household	1.355	0.35	-2.161	1.43
	Years of education of head of household	-137.6	0.86	20.43	1.57
	Household labor days per hectare	-0.0826	0.34	**3.200****	**3.02**
	Hired labor days per hectare	2.543	1.30	**12.61****	**3.00**
	Household owned/rented ox, ox plough, ox planter, or ox cart=1	1330.8	1.29	171.0	1.49
	Household received advice from any source=1	-204.0	0.60	130.9	0.71
	Household sold maize=1	-165.5	0.29	**575.2*****	**5.51**
	Constant	4253.1	1.38	502.4	0.78

*** p<0.01, ** p<0.05, * p<0.1 | Values in bold represent significant variables. Shading indicates significance in one model but not in the other.

Rice Area Harvested, Tanzania