Are Small Farms Really more Productive than Large Farms?*

Fernando Aragon
Simon Fraser University

Diego Restuccia
University of Toronto
and NBER

Juan Pablo Rud
Royal Holloway, University of London
and Institute for Fiscal Studies

June 2019

Abstract

We revisit the long-standing empirical evidence of an inverse relationship between farm size and productivity using rich micro data from Uganda. We show that farm size is negatively related with yields (output per hectare), as commonly found in the literature, but positively related with farm productivity (a farm-specific component of total factor productivity). These conflicting results do not arise because of omitted variables such as land quality, measurement error in output or inputs, or specification issues. Instead, we reconcile the findings emphasizing decreasing returns to scale in farm production and farm-specific distortions. We exploit regional variation in land tenure regimes in Uganda in evaluating the role of farm-specific distortions. Our findings point to the limited value of yields (or land productivity) in establishing the size-productivity relationship. More generally, we highlight farm size as an ineffective instrument for policy implementation since size is deeply confounded by distortions in developing countries.

JEL classification: O4, O5, O11, O14, E01, E13.
Keywords: Farm size, productivity, yields, land markets, agriculture, policy.

*For useful comments, we are grateful to seminar participants at several institutions and conferences. All errors are our own. Restuccia gratefully acknowledges the support from the Social Sciences and Humanities Research Council of Canada and the Canada Research Chairs program. *Contact Information:* Aragon: Department of Economics, Simon Fraser University, Burnaby, BC, V5A 1S6, Canada (e-mail: faragons@sfu.ca); Restuccia: Department of Economics, University of Toronto, 150 St. George Street, Toronto, ON M5S 3G7, Canada (e-mail: diego.restuccia@utoronto.ca); Rud: Department of Economics, Royal Holloway, University of London, Egham, Surrey TW20 0EX, UK (e-mail: juan.rud@rhul.ac.uk).

1 Introduction

An important and established literature has documented a robust inverse relationship between yields (i.e., output per unit of land) and farm size. The implication is that if the yield proxies for productivity, a common presumption in the literature, then small farms are more productive than large farms (Barrett, 1996, Barrett et al., 2010, Berry et al., 1979). This interpretation of the inverse yield-size relationship has had a profound effect on agricultural policy whereby small landholder agriculture is considered a cornerstone of microeconomic development (Collier and Dercon, 2014). This view contrasts sharply with the macroeconomic evidence whereby farming in poor countries operates predominantly at small scale and feature much lower agricultural productivity, compared to rich countries where the predominant unit of production is large scale farming (Adamopoulos and Restuccia, 2014).

In this paper, we argue that using yields as a measure of productivity to establish the size-productivity relationship is not informative for agricultural policies and can lead to erroneous recommendations. This happens because yields pick up not only farm productivity, but also decreasing returns to scale (DRS) and market distortions. These issues can result into qualitatively different estimates of the farm size-productivity relationship.

We illustrate this point using rich micro data from Ugandan farmers. We construct two alternative measures of productivity: (1) yields, as is standard in the literature and (2) farm productivity, exploiting the detailed information from the micro data. Specifically, farm productivity is constructed by estimating a farm-level production function and corresponds to the farm-specific component of total factor productivity (TFP). Using these two alternative measures, we evaluate the farm size-productivity relationship. We find that the results are highly sensitive to the measure of productivity we use. We find a negative relationship between yields and farm size consistent with the broad findings documented in the literature. Interestingly, the quantitative magnitude of the relationship for Uganda is quite close to that reported for other countries. But when using farm productivity instead of yields, we find a *positive* relationship with farm size. These conflicting results do not appear to be generated as an statistical artifact due to omitted soil characteristics or measurement errors on farm size or output.

Instead, our results reflect a more profound limitation of yields as a measure of productivity by farm size. Since yields are a measure of land productivity, yields pick up farm productivity plus deviations from constant returns to scale (CRS) and market distortions, which get confounded when comparing farms by farm size. In practice, these considerations imply that estimates of the size-productivity relationship using yields are inconsistent when the farm technology exhibits DRS or in the presence of size-dependent distortions. These are plausible conditions in most applications in developing countries.

We evaluate the validity of our interpretation of the results in two ways. First, we revisit our estimates of the size-productivity relationship and show that, after correcting for DRS and market distortions (using the input ratio as a proxy), the negative relationship turn into a positive relationship between yields and farm size. Second, we exploit variation in land tenure regimes to examine in more detail the role of market distortions. Uganda has two broad types of tenure regimes: non-customary and customary systems. Customary systems are based on communal ownership, are perceive as less secure and face higher transaction costs. We use the type of land tenure as a measure for the extent of market distortions. The key idea is that in places with modern, non-customary, land (such as Western and Central regions), markets distortions would be smaller and thus the yield-farm size relationship would be less biased. We start by showing that factor misallocation (measured by the correlation between farm productivity and input use) is smaller in places with non-customary land. Then, we find that the yield-farm size relationship becomes less negative in areas with non-customary property rights. We interpret this finding as suggestive evidence that market distortions play a role on driving the negative yield-farm size result.

Our paper is not the first to raise concerns on the use of yields as a measure of productivity when evaluating the size-productivity relationship. Early reviews on the size-productivity relationship have emphasized the use of yields as “flawed by methodological shortcomings” (Binswanger et al., 1995, p. 2706). More recently, Helfand et al. (2018) argue that partial measures of productivity, such as yields, may be inappropriate and advocate for using instead total factor productivity. Using data from Brazil, Helfand et al. (2018) also show that the sign of the size-productivity relationship depends of the measure used, and suggest that deviations from CRS could be a possible source of discrepancy.

The controbutio of our paper to this debate is twofold. First, we clarify why yields are not informative of the size-productivity relationship, and may lead to erroneous policy recommendations. In particular, we point to deviations from CRS and size-dependent market distortions as potential sources. Second, we provide empirical evidence on the relevance of this problem, and examine the quantitative importance of these factors (i.e., DRS and market distortions) on explaining the negative relationship.

More importantly, our results also point out a broader limitation of the size-productivity literature. While the inverse size-productivity relationship provides a tractable mechanism for policy implementation—since farm size can be used as a proxy for productivity and thus employed to target farmers and enhance economic efficiency—our results highlight that farm size is an ineffective instrument for policy since productivity by farm size is deeply confounded by distortions. This happens not only because the sign of the size-productivity relationship may be wrongly estimated—as it is the case when using yields—but also because distortions imply substantial dispersion in productivity across farms of the same size, rendering farm size a poor proxy of productivity. We illustrate this point by documenting the dispersion in productivity (using both yields and farm productivity) across farms within farm-size categories (for example, 0 to 1 ha, etc.) and showing that this dispersion is large for each farm size class, in some cases larger than the dispersion of productivity across the entire sample of farms. Moreover, the ratio of farm productivity between the 90th and 10th percentiles of farms is a factor of more than 8-fold, whereas the ratio of average farm productivity between the largest and smallest farm size classes is a factor of only 2.3-fold (the ratio of average yields between the smallest and largest farm size categories is only a factor of 1.4-fold).

The paper is organized as follows. In Section 2, we provide some background on the farm size-productivity relationship and its importance to guide economic policy. Section 3 presents the empirical evidence from Uganda and show that, using alternative measures of productivity, produces different estimates of the farm size-productivity relationship. Section 4 examines, theoretically and empirically, the reasons for these conflicting results. We conclude in Section 5.

2 Background

The study of the relationship between farm size and productivity occupies a central place in the agrarian and development economics literature (Barrett et al., 2010). An important finding in this literature is the inverse relationship between farm size and yields (output per unit of land). This result has been documented in several countries in Asia, Africa, and Latin-America (Barrett, 1996, Barrett et al., 2010, Berry et al., 1979) and has been interpreted as evidence that small farms are more productive.

There are several explanations in the literature for why small farms may be more productive. For instance, small farms may be able to solve a contractual problem (hidden effort) of farm workers (Eswaran and Kotwal, 1986, Feder, 1985). A complementary explanation is that small farms may be more productive due to selection: low-productivity farmers with small landholdings may be more likely to hit a minimum consumption threshold, and leave agriculture, than similar farmers with larger farms. As a result, in average, small farms would be more productive (Assuncao and Ghatak, 2003). A recent explanation emphasizes a behavioral phenomenon: farmers may put more effort on the edges of a plot. This "edge" effect would be proportionally larger in smaller plots—which have relatively more area on the edges—and thus explain why small farms are more productive (Bevis and Barrett, 2016).

The interest on the farm size-productivity relationship stems, in part, from its profound normative implication: if small farms are indeed more productive, then land redistribution from large to small farms can increase agricultural productivity and food availability (Barrett et al., 2010, Collier and Dercon, 2014). In addition, the farm size-productivity relationship provides a tractable approach to policy implementation. Since farm size is easy observable whereas productivity is difficult to assess in real time, policies are often implemented based on size.

To fix ideas, consider a standard model of farm size and input allocation. The framework is based on Adamopoulos and Restuccia (2014), building on previous work from Lucas Jr (1978) and Hopenhayn (1992). There are n heterogeneous farmers producing a single homogeneous good according to the following production function:

$$y_i = s_i A (T_i^{\alpha} L_i^{1-\alpha})^{\gamma}, \tag{1}$$

where T_i and L_i stand for the amounts of land and labor used by farmer i. Total factor productivity is equal to $s_i A$, where A is a common productivity shock, such as weather, and s_i is a farm-specific output shifter, such as farming ability or entrepreneurship. Henceforth, we call s_i farm productivity.

This framework provides a simple way to describe the efficient, first best, allocation of land and labor across farmers. Consider a static allocation in which the set of farmers and distribution of productivity are given, and the economy has fixed endowments of land (T^e) and labor (L^e). In this context, the efficient allocation maximizes aggregate output and solves the following social planner's problem:

$$\max_{\{T_i, L_i\}} \quad \sum_i s_i A(T_i^{\alpha} L_i^{1-\alpha})^{\gamma},$$
$$\text{subject to} \quad \sum_i T_i = T^e, \ \sum_i L_i = L^e.$$

The Pareto efficient allocation equates the marginal product of land and labor across farmers. By the first welfare theorem, this solution also corresponds to the allocation in a competitive market equilibrium with no distortions. Letting $z_i \equiv s_i^{1/(1-\gamma)}$, we can characterize the efficient allocations as:

$$T_i^* = \frac{z_i}{\sum z_i} T^e, \ L_i^* = \frac{z_i}{\sum z_i} L^e.$$

The main insight is that an efficient allocation requires resources to be *proportional* to farm productivity s_i, that is, more productive farmers should operate more land and labor. It follows that if the relationship between farm size and farm productivity is indeed negative, then policies that redistribute land towards small landholders would enhance economic efficiency and increase aggregate productivity. The framework also suggests that an approach to assess the extent of factor misallocation in an economy involves comparing the relationship between farm size and farm productivity in the actual data to the efficient benchmark. This source of economic inefficiency has attracted attention as a quantitatively relevant determinant of income differences across countries (Hsieh and Klenow, 2009, Restuccia et al., 2008).

An important issue is that the evidence that small farms are more productive is based on estimates of the relationship between farm size and yields (or land productivity) instead of measures of farm productivity s_i. We argue below that, in many applications, this choice of variable can result in misleading estimates of the size-productivity relationship and thus lead to erroneous policy recommendations. Moreover, we argue that because of distortions and other market imperfections farm size hides substantial heterogeneity in farm productivity, diminishing the effectiveness of size-related policy implementations.

3 Empirical evidence

We revisit the evidence that small farms are more productive than large farms using microdata from Ugandan households. We start by replicating the inverse relationship between yields and farm size, as in the existing literature. Then, we re-examine this result with the same data but using estimates of farm productivity s_i.

3.1 Data

We use data from the Uganda Panel National Survey (UPNS), a household-level panel dataset collected with support from the World Bank, as part of the LSMS-ISA project. This survey is representative at the urban/rural and regional level and covers the entire country. We use the four available rounds: 2009-10, 2010-11, 2011-12, and 2013-14. Every round collects agricultural information for each of the two cropping seasons (i.e. January to June and July to December). We focus on the farm as the production unit so our unit of observation is a household farm i in period t, where t refers to a season-round pair. A farmer may operate one or several parcels or plots of land and hence we aggregate any information at the parcel level to the household-farm level. Focusing on the farm as the production unit is important in measuring productivity. Our dataset contains a panel of around 3,400 farming households observed, on average, for four periods. Figure A.1 in the Appendix displays the map of Uganda and sample coverage.

Output and inputs We construct measures of agricultural output and input use (i.e., land and labor) for each farm in a given period. To measure real agricultural output at the farm level, we construct a Laspeyres index of production that aggregates the quantity produced of each crop by the household farm using proxies of prices in 2009 as weights. We use unit values as proxies of prices. To calculate these proxies, we divide the value of sales by the quantity sold of each crop. Then, we obtain the median unit value of each crop at the national level.

We measure area of land cultivated by adding up the size of parcels planted by the household. Similar to previous studies, we use this variable as our main measure of land use and farm size. In addition, we obtain measures of available land from self-reported information and from GPS data. The available land corresponds to all the parcels of land the farmer has access to either because the farmer owns the land or has user rights, for instance, due to rental agreements. We use these variables as measures of land endowment, and as alternative proxies of farm size.

Our measure of labor input is the total number of person-days used in the farm. The survey distinguishes between work done by household members and by hired workers. We use this information to construct measures of family and hired labor.

Other variables The survey also provides information on agricultural practices (such as use of fertilizers, pesticides, or intercropping), soil characteristics and land tenure regimes. Regarding soil characteristics, the survey asks farmers to classify each parcel according to soil type (sand loan, sandy clay loan, back clay or other), quality (good, fair, or poor) and topography (hilly, flat, gentle slope, valley or other). We aggregate the parcel-level indicators to the farm level to obtain a share of farm land in each category.

Similarly, we obtain indicators of the share of land (at the farm and district level) under different tenure regimes. We distinguish two types of tenure regimes: customary and non-customary. Non-customary tenure regime includes freehold, leasehold, and Mailo.[1] Later, we use these variables as a measure of property rights to assess the role of land market imperfections.

[1] The Mailo tenure system is a form of leasehold in which land owners hold their land in perpetuity and have similar rights to freeholders, while tenants have security of occupancy as in common-law arrangements (sometimes backed by a certificate). Tenant can only be removed if the land is unattended for at least three years (Coldham, 2000).

We complement the household survey with weather data: temperature and precipitation. These variables are relevant determinants of agricultural productivity (cite). We use high-frequency satellite imagery and gridded data to obtain measures of cumulative exposure to heat and water. For temperature, we use the MOD11C1 product provided by NASA. The satellite data provides daily estimates of land surface temperature (LST). Precipitation data comes from the Climate Hazards Group InfraRed Precipitation with Station data (CHIRPS) product (Funk et al., 2015). We combine the weather and survey data using the location of the sub-county (n=967) of residence of the household. Our approach to model exposure to weather is similar to previous work (Schlenker and Roberts, 2006, 2009). In particular, we obtain average precipitation, degree days and harmful degree days during the last cropping season for each farmer. Degree days (DD) measures the cumulative exposure to temperatures between 8℃ and 26℃ while harmful degree days (HDD) captures exposure to temperatures above 26℃. The inclusion of HDD allows for potentially different, non-linear, effects of extreme heat.

Table 1 presents summary statistics of our main variables. There are several relevant observations. First, farmers have small scale operations (the average cultivated area is 2.3 hectares). Second, farmers use practices akin to subsistence agriculture such as inter-cropping (i.e, cultivation of several crops in the same plot), reliance on domestic instead of hired labor, and limited use of fertilizers, pesticides, and capital inputs. Finally, there is substantial variation in land property rights: around 46% of the land is held under non-customary regimes (like freehold, leasehold, and Mailo) while the rest is held under customary, communal, property rights.

Measures of productivity We construct two alternative measures of productivity: land productivity (or yield) and farm productivity.[2] First, we calculate yields (Y/T) by dividing

[2] We refer to our measure of real farm output per unit of operated land as land productivity or yield interchangeably.

Table 1: Summary statistics (UPNS 2009-2014)

Variable	Mean	Std. Dev.
HoH age	47.2	15.2
HoH can read and write	0.657	0.475
HoH is female	0.222	0.416
Household size	6.1	2.9
Total output (in 2009 Ush, 000s)	2854.4	6118.0
Yields (output per ha.)	5013.6	7510.0
Land cultivated (has)	2.300	2.136
Land available (has)	4.247	10.713
Land available - GPS (has)	2.606	17.015
Total labor (person-day)	125.5	97.0
Hired labor (person-day)	14.1	170.6
Domestic labor (person-day)	124.0	119.4
% farm land using org. fertilizer	0.071	0.231
% farm land using intercropping	0.565	0.429
% farm land non-customary tenure	0.459	0.487
Degree days (℃)	15.1	1.8
Harmful degree days (℃)	1.0	1.0
Precipitation (mm/month)	105.8	50.7

Notes: Sample restricted to farming households. HoH = Head of household. Non-customary land tenure includes freehold, leasehold, and Mailo.

real farm agricultural output, at 2009 prices, by the area of land cultivated. This variable is similar to measures of crop yields used in previous work. The key distinction is that we use the value of total agricultural farm output (using time invariant and common prices across farms) instead of the quantity produced of a single crop. This distinction arises because of our focus on the farm rather than the plot as the main production unit and the associated multi- and inter-cropping: farmers usually cultivate several crops, sometimes even in the same plot. These features make it difficult to attribute inputs (land or labor) to individual crops.

Second, we obtain estimates of farm productivity s_i. To do so, we estimate the following production function $Y_{ijt} = s_i A_{ijt}(T_{it}^{\alpha} L_{it}^{1-\alpha})^{\gamma}$, where the unit of observation is a household farm i, in location j, and period (season-year) t. We assume that the common productivity shock is $A_{ijt} = \exp(\delta \text{weather}_{jt} + \eta_{jt} + \epsilon_{ijt})$ where η_{jt} is a region-season-year fixed effect, weather_{jt} is a set of temperature and precipitation variables, and ϵ_{ijt} is the error term. Taking logs, we obtain:

$$\ln Y_{ijt} = \ln s_i + \alpha\gamma \ln T_{it} + (1-\alpha)\gamma \ln L_{it} + \delta \text{weather}_{jt} + \eta_{jt} + \epsilon_{ijt}. \tag{2}$$

We estimate equation (2) using panel data methods with household fixed effects. Our preferred specification is a Cobb-Douglas in land and labor with same parameters for all regions. However, we check the robustness of our results using estimates of farm productivity s_i obtained from a translog production function, a Cobb-Douglas production with heterogeneous parameters (α, γ) by region, and using input endowments (available land and household size) as instruments for land and labor. The estimated production function parameters are $\hat{\alpha} = 0.531$ and $\hat{\gamma}$ =0.715, which are close to the values calibrated in the context of similar economies such as Restuccia and Santaeulalia-Llopis (2017) for Malawi and Adamopoulos et al. (2017) for China. Table A.1 in the Appendix present detailed results of the production function estimation. We use the estimated fixed effects of our baseline specification as measures of $\ln s_i$, the log of farm productivity. Figure A.2 in the Appendix reports the resulting distribution of the estimated household-farm fixed effects.

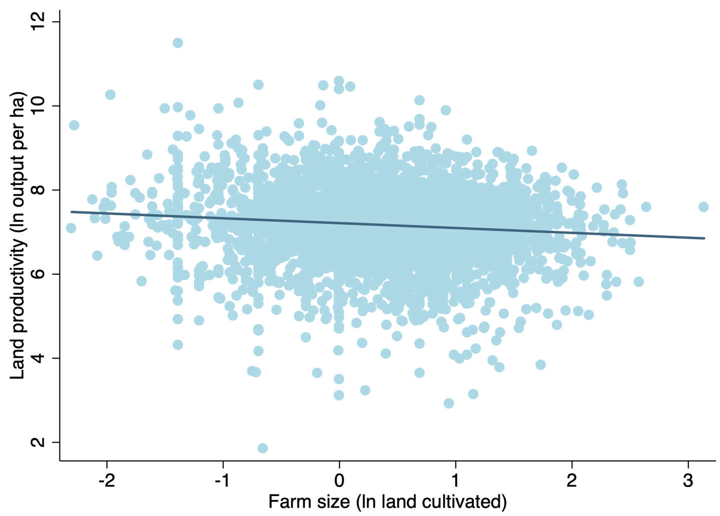
We note that there is a strong positive correlation between land productivity and farm productivity of 0.86. (See Figure A.3 in the Appendix for a documentation of the relationship between our two measures of productivity.) Despite these similarities, we show below that they produce qualitatively different estimates of the farm size-productivity relationship. Furthermore, the positive relationship between yields and farm productivity is prima facie evidence of misallocation and market imperfections. To see this, note that under the Cobb-Douglas assumption, Y/T is proportional to the marginal product of land. In the efficient allocation, this marginal product is equalized across farmers, regardless of their productivity. Thus, yields should be uncorrelated to s_i.

3.2 Conflicting findings depending on the measure of productivity

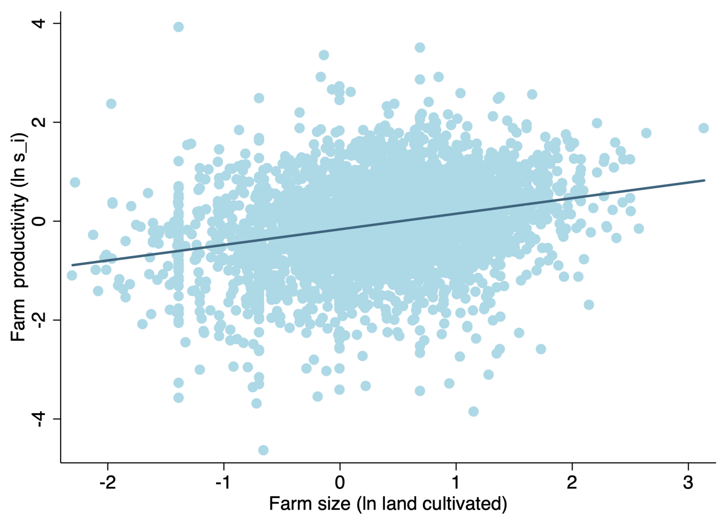
Figure 1 displays the relationship between the log of cultivated area, our baseline measure of farm size, and the two measures of productivity. A important observation is that the relationship is qualitatively different depending of the measure of productivity used. Using yields (panel A), we observe a negative relationship. This finding is consistent with previous results of an inverse farm size-productivity relationship. However, when using farm productivity (panel B), the relationship is *positive*.

Table 2 presents a formal analysis of the inverse relationship between yields and farm size. We employ two specifications commonly used in the farm size-productivity literature: the yield approach and the production function approach (Ali et al., 2015, Assunção and Braido, 2007, Barrett et al., 2010, Carter, 1984). The yield approach regresses log of yields on log of land cultivated and includes a host of control variables. We include a rich set of soil characteristics, weather controls (temperature and precipitation), farmer characteristics (such as age, literacy and gender), as well as region-by-period and district fixed effects. The production function approach adds to the previous specification the log of the labor-land ratio. Assuming a Cobb-Douglas technology with constant returns to scale, this specification is equivalent to estimating the production function. We present results using both specifications and varying the set of covariates. We also check the robustness of our results to using (self-reported) available land as measure of farm size, and to collapsing the panel data by taking the average for each household (see Table A.2 in the Appendix). In all cases, we find a negative and significant relationship between farm size and yields. Interestingly, the

Figure 1: Farm size and productivity



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



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

estimated coefficient (around -0.25 in our preferred specification) is similar in magnitude to previous estimates using data from other countries (Barrett et al., 2010, Desiere and Jolliffe, 2018).

We replicate the analysis using farm productivity ($\ln s_i$) instead of yields and report the results in Table 3. The results confirm the conflicting patterns observed in Figure 1: there is a robust and significant *positive* relationship between farm size and farm productivity (see results in columns 1 and 2 for specifications without and with controls). One potential concern with these results is that we are artificially obtaining statistically significant results by duplicating the time-invariant measure of farm productivity in the panel data, however, this turns out not to be an issue as we obtain qualitatively similar results collapsing the panel data at household level (column 3). Our baseline specification uses estimates of s_i obtained from a Cobb-Douglas production function. However, this choice of functional form does not seem to be driving the results. We obtain similar results using estimates of s_i obtained with more flexible specifications, such as translog production function, a Cobb-Douglas with heterogeneous parameters by region, or estimating the production function using endowments as instruments for input used (columns 4 to 6). Our findings are also robust to using land available as measure of farm size (see Table A.2 in the Appendix.)

3.3 A statistical artifact?

Existing work suggests that the inverse yield-farm size relationship may be driven by omitted variables, e.g. soil quality (Benjamin, 1995), or systematic measurement error (Carletto et al., 2013, Desiere and Jolliffe, 2018, Gourlay et al., 2017). This error arises if small farmers over-report output or under-report land. The measurement error could generate the inverse relationship between yields and farm size, even if the actual relation is insignificant. A relevant concern is that the pattern of results we observe may be a statistical artifact of these identification problems.

Table 2: Yields and farm size

	Outcome variable: ln(output per ha)					
	Yield approach			Production function approach		
	(1)	(2)	(3)	(4)	(5)	(6)
ln(land cultivated)	-0.239***	-0.257***	-0.487***	-0.035**	-0.064***	-0.295***
	(0.015)	(0.014)	(0.019)	(0.016)	(0.016)	(0.021)
ln(labor/land)				0.231***	0.220***	0.175***
				(0.014)	(0.014)	(0.015)
Controls	No	Yes	Yes	No	Yes	Yes
Household FE	No	No	Yes	No	No	Yes
No. obs.	16,063	14,577	15,787	15,806	14,334	15,532
R-squared	0.029	0.175	0.110	0.087	0.217	0.145

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%. All regressions (except in column 1) include district and region-by-year fixed effects as well as soil, farmer and weather controls. Soil controls= % of farm land of different types, quality and topography. Farmer controls = age, literacy, gender, ethnic group. Weather controls: DD, HDD and log of precipitation. Columns 3 and 6 also include household fixed effects.

Table 3: Farm productivity and farm size

	Outcome variable = farm productivity ($\ln s_i$)					
	(1)	(2)	(3)	(4)	(5)	(6)
ln(land cultivated)	0.213***	0.186***	0.296***	0.200***	0.164***	0.193***
	(0.012)	(0.011)	(0.022)	(0.011)	(0.011)	(0.011)
Controls	No	Yes	Yes	Yes	Yes	Yes
Prod. function used to estimate s_i	CD	CD	CD	CD by region	CD + IV	Translog
No. obs.	16,748	15,361	3,249	15,361	15,279	15,361
R-squared	0.060	0.336	0.301	0.341	0.820	0.381

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%. All regressions (except in column 1) include soil and farmer controls similar to Table 2, as well as district fixed effects. CD=Cobb-Douglas in land and labor inputs. Column 3 uses a cross-section of farmers obtained by collapsing the panel data at household level. Column 4 estimates s_i using a flexible CD specification with different parameters by region, column 5 uses a CD specification that instruments input use with input endowments (land available and household size), while column 6 uses a translog production function.

We examine this possible explanation in several ways. First, our regressions control for a rich set of soil characteristics, and are robust to including district and household fixed effects. These findings weaken the argument that our results are affected by omitted variables. Second, we replicate our baseline results using, as proxies of farm size, the area of available land measured using a GPS device. Arguably, this variable is less prone to have a systematic measurement error than self-reported land. The results are, however, qualitatively similar (see columns 1 and 2 in Table 4).[3]

Finally, we examine the role of systematic measurement error in self-reported output. We do not have alternative measures of output, such as cross-cuttings. However, under the assumption that the measurement error is correlated with farm size, we can control for it when estimating the relationship between farm productivity and farm size. In particular, we

[3]Note that we loose some observations because GPS measures are only available for a random sub-sample of farmers.

modify equation (2) by assuming that $\epsilon_{ijt} = v_{ijt} + M(T_i)$, i.e., there is systematic measurement error which is a function of farm size. Note that omitting $M(T_i)$ as a regressor would create an endogeneity problem. Empirically, we address this issue by approximating M with a 4^{th} degree polynomial of the GPS measures of available land and including these variables as additional regressors when estimating s_i. Table 4, column 3, shows that the positive relationship between farm productivity and farm size is robust to using these alternative estimates of s_i.

Taken together, we interpret these results as evidence that the opposite findings on the farm size-productivity relationship documented in Tables 2 and 3 are not due to omitted variables or systematic measurement error. So, what explains these different results?

Table 4: Farm productivity and farm size

	ln(output per ha.) using GPS measure	farm productivity	
	(1)	(2)	(3)
ln(land available) GPS measure	-0.629***	0.127***	0.133***
	(0.016)	(0.010)	(0.010)
Prod. function		CD	CD + land polynomial
No. obs.	10,256	11,146	11,146
R-squared	0.392	0.358	0.399

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%. Column 1 same controls as column 2 in Table 2. Columns 2 and 3 use same controls as column 1 in Table 3.

4 What explains the different results?

We argue that, in many applications, using yields as a measure of productivity is not informative of the farm size-productivity relationship. This occurs because yields pick up not only farm productivity, but also market distortions and decreasing returns to scale. These issues can lead, as in the case of Uganda, to wrongly inferring a negative relationship.

To illustrate this, consider a researcher who wants to examine the farm size-productivity relationship. The "true" model the researcher wants to estimate is:

$$\ln s_i = \beta \ln T + \epsilon, \tag{3}$$

where as before s_i is farm productivity, T is land size, and ϵ is an error term. Assume that farms have a production function as described in equation (1), that is $y_i = s_i A \left(T_i^{\alpha} L_i^{1-\alpha}\right)^{\gamma}$, and face potentially imperfect markets. Following Hsieh and Klenow (2009), we model market distortions as 'wedges' or taxes on input prices. Without loss of generality, we assume that the price of labor is w while the price of land is $r(1+\tau_i)$. Note that τ_i has a broad interpretation. It can be interpreted as subsidies or taxes, but also as any other market imperfection or institutional feature that distorts effective relative input prices. We allow for these distortions to be (potentially) different across farms. The wedge τ_i measures the relative distortion in input markets and so we are implicitly normalizing the distortion in labor prices equal to one. The special case of efficient markets occurs when $\tau_i = 0$.

Profit maximization implies that farmer i chooses the following input ratio:

$$\frac{L_i}{T_i} = \frac{1-\alpha}{\alpha}\frac{r}{w}(1+\tau_i). \tag{4}$$

Using this result and taking logs, we can re-write yields $\frac{Y_i}{T_i} = s_i \frac{\left(T_i^{\alpha} L_i^{1-\alpha}\right)^{\gamma}}{T}$ as:

$$\ln \frac{Y_i}{T_i} = \ln s_i + \gamma(1-\alpha)\ln(1+\tau_i) + (\gamma - 1)\ln T_i + c \tag{5}$$

where c is a constant, which is a function of common prices and parameters (w, r, α, γ) and the common productivity shock (A).

The researcher uses yields as a proxy for farm productivity s_i. This estimated model is equivalent to the yield approach regression used in the farm size-productivity literature. Replacing (3) into (5), we can see that the estimated model is:

$$\ln \frac{Y_i}{T_i} = c + \beta \ln T + \mu \tag{6}$$

where the error term is: $\mu = \gamma(1-\alpha)\ln(1+\tau_i) + (\gamma - 1)\ln T_i + \epsilon$. Equation (6) highlights two reasons why, in general, using yields, can lead to inconsistent (wrong) estimates of the relationship between productivity and farm size: (1) decreasing returns to scale, or (2) size-dependent market distortions. These are plausible conditions in many applications, especially in the context of subsistence farmers in developing countries.[4] In either case, the error term μ would be, by construction, correlated to farm size and OLS estimates of β would be inconsistent. OLS estimates of β would be consistent only in very special cases such as (1) efficient markets ($\tau_i = 0$) and constant returns to scale (CRS), or (2) CRS and distortions independent of farm size (i.e., τ_i uncorrelated to T.

This problem cannot be solved by adding better controls of soil quality or other determinants of farm productivity, nor by reducing measurement error on land or output. Similarly, if the technology exhibits DRS, the problem would persist even after using instruments or even randomizing farm size.

The source of the problem is more profound: it arises from using yields, a proxy of land productivity, instead of measures of productivity of the production unit, i.e. farm productivity. Yields can be affected by DRS and size-dependent market distortions. As we

[4]These two factors were among the first explanations of the inverse farm size-yield relationship (Bardhan, 1973, Carter, 1984, Sen, 1962). However, they have been ruled out as leading explanations for at least two reasons. First, early studies find evidence of constant returns to scale in agriculture Berry et al. (1979). Second, several studies document an inverse size-yield relationship among plots of the same household (Assunção and Braido, 2007, Desiere and Jolliffe, 2018). Note, however, that our production function estimates reject CRS. In addition, the inverse relationship at plot level does not proof that market distortions are not relevant, only that they are not the *only* driver of the inverse relationship.

will see below, when correcting for these issues, we can reverse the original results and obtain a positive relationship between yields and farm size.

4.1 Correcting for DRS and market distortions

Equation (4) suggests that the observed labor-land ratio is proportional to the market distortion. Using the input ratio as a proxy for $(1+\tau_i)$ and the “true" farm-size productivity relationship (equation 3), we can rewrite expression (5) as follows:

$$\ln\frac{Y_i}{T_i} = \text{constant} + (\beta+\gamma-1)\ln T + \gamma(1-\alpha)\ln(\frac{L_i}{T_i}) + \epsilon. \tag{7}$$

This expression suggests using a specification similar to the “production function approach” used in the existing literature. That is, regressing yields on farm size and the input ratio. A key distinction, however, is that we do not impose CRS. Instead, we use a value of $\hat{\gamma} = 0.711$ obtained from estimating the production function (see column 1 in Table A.1). This is relevant, because the estimate associated with farm size is $\beta+\gamma-1$. Thus, to recover β, the farm-size productivity relationship, we also need to account for possible deviations from CRS. We do so by subtracting $(\hat{\gamma}-1)$ from the estimates associated with farm size.

Table 5 presents the estimates of equation (7) using two alternative measures of farm size, T_i: (self-reported) area cultivated and GPS measures of available land. We start by replicating the “yield approach" (column 1 and 4) and then gradually adding the input ratio, our proxy for market distortions, (columns 2 and 5) and relax the CRS assumption (columns 3 and 6). The main result is that the initially negative estimate of the slope coefficient between yields and farm size becomes less negative after correcting for market imperfections, and eventually becomes positive when relaxing the assumption of CRS.

These results support our argument that, in many applications, yields are not useful to examine the farm size-productivity relationship. Yields pickup not only farm productivity

Table 5: Correcting for DRS and market distortions

	Outcome variable: ln(Y/T)					
	(1)	(2)	(3)	(4)	(5)	(6)
β: ln(T)	-0.257***	-0.064***	0.225***	-0.629***	-0.181***	0.108***
	(0.014)	(0.016)	(0.016)	(0.016)	(0.020)	(0.020)
$\gamma(1-\alpha)$: ln(L/T)		0.390***	0.390***		0.579***	0.579***
		(0.017)	(0.017)		(0.019)	(0.019)
Adding input ratio	No	Yes	Yes	No	Yes	Yes
Relaxing CRS	No	No	Yes	No	No	Yes
Measure of T_i	area cultivated (self-reported)			GPS measure of available land		
Implied α		0.610	0.451		0.430	0.186
Assumed γ	1.000	1.000	0.711	1.000	1.000	0.711
No. obs.	14,577	14,334	14,334	10,256	10,060	10,060
R-squared	0.175	0.217	0.195	0.392	0.469	0.343

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%. All regressions include district, region-by-year fixed effects and soil, weather and farmer controls as column 2 in Table 2. $\hat{\gamma} = 0.711$ obtained from Column 1 Table A.1. Implied α obtained from estimates associated with ln(L/T) (second row).

but also size-dependent distortions and DRS. These issues are quantitatively important and, as shown in the case of Uganda, can lead to substantially different conclusions.

4.2 Using land tenure regimes as proxies of market distortions

Our previous results use the input ratio L/T as a proxy for market distortions. However, the validity of this proxy depends on the functional form assumption of the production function. For example, consider an alternative CES specification $f(T_i, L_i) = [A_i T^\rho + B_i L^\rho]^{\frac{\gamma}{\rho}}$ where A_i and B_i are input-specific productivity shifters that can vary by farmer. In that case, the land-labor ratio would be $\frac{B_i}{A_i}\frac{r}{w}(1+\tau_i)$. Thus, the input ratio would pickup not only the market distortion but also differences in input-specific productivity $\frac{B_i}{A_i}$. Similarly, it is also possible that land market distortions are highly correlated with other input distortions, making the land-labor ratio less responsive to distortions and hence making it a poor proxy of land market distortions. This is indeed what the evidence from different contexts in agriculture suggests (Restuccia and Santaeulalia 2017, Adamopoulos et al 2017, Chen et al 2017).[5] As an alternative approach, we exploit variation in land tenure regimes as a proxy of market distortions. This approach is motivated by the Coase theorem, and existing evidence that well-defined property rights play an important role on improving allocative efficiency (CITE).

There are four types of land tenure in Uganda: freehold, leasehold, Mailo (form of freehold), and customary land. The first three tenure systems offer some degree of formal, secure, property rights. In contrast, customary systems are based on communal ownership, are perceived as less secure and may face higher transaction costs due to lack of formal land registries and community approval requirements (Coldham, 2000, Place and Otsuka, 2002). These differences in land tenure characteristics seem to matter for economic activity. For instance, Place and Otsuka (2002) find that customary land is associated with less agricultural investment. In our data, we also observe that in regions with more prevalent use of non-customary tenure systems around 47% of land holdings have been marketed, i.e., acquired through purchase or rented. In contrast, in regions with more customary land, this figure is much lower, around 27%.

These tenure systems are spatially concentrated (see Figure A.4 in the Appendix). Customary land is dominant in the Northern and Eastern regions, where more than 90% of land holdings are under this regime. In contrast, non-customary systems are mostly found in the Western and Central regions. In these regions, less than 7% of land is held under customary systems. In our empirical analysis, we use regional indicators as the main proxies for the quality of property rights and development of land markets.[6]

[5]The results in Hsieh and Klenow (2009) for the manufacturing sector in China and India also indicate that distortions to the capital-labor ratio account for a small fraction of the dispersion in overall distortions, implying that input wedges are highly correlated between them or have the same underlying source.

[6]Results are robust to using a more granular measure, such as share of non-customary land in the district

We start by assessing whether land tenure differences matter for factor misallocation. To do so, we estimate the relation between input use (land and labor) and farm productivity ($\ln s_i$). As discussed in section 2, in an efficient allocation, these variables should be positively correlated. We allow for different values by type of land tenure by including an interaction term with an indicator of being in the Western/Central region.

Table 6, columns 1 and 2, display the results. Note that the regressions in this table use the GPS measure of available land as a proxy for farm size. We use this variable to reduce concerns of measurement error in self-reported data. We find a positive relationship which is larger in regions with less customary land (i.e., Western and Central). These results are consistent with modern property rights improving allocative efficiency. Note, however, that these results suggest a substantial amount of misallocation in all regions. Assuming a Cobb-Douglas technology in land and labor, the estimated relationship between input use and farm productivity in the case with no distortions should be equal to $\frac{1}{1-\gamma}$. Given our estimate of $\hat{\gamma} = 0.715$, the implied slope is around 3.5. In contrast, the estimated slope in the data is quite small: between 0.17 to 0.48. These results echo similar findings in other contexts (Adamopoulos and Restuccia, 2019, Adamopoulos et al., 2017, Restuccia and Santaeulalia-Llopis, 2017).

Next, we examine whether the inverse yield-farm size relationship is affected by land tenure. Columns 3 and 4 in Table 6 replicate the inverse relationship (using both the yield and production function approaches) and show that it becomes less negative in regions with better property rights. Note that in both cases, we are not correcting for DRS, but instead keep the assumption of CRS as in the existing literature. This evidence is consistent with our interpretation that the negative relationship between yields and farm size reflects, in part, market distortions.[7]

(see Table A.4 in the Appendix).

[7] Several studies find that the inverse size-productivity relationship persists even when comparing plots within the same household (Assunção and Braido, 2007, Desiere and Jolliffe, 2018). These results have been interpreted as evidence that imperfect input markets cannot explain the inverse farm size-yield relationship.

Table 6: Heterogeneity by region

	Assessing misallocation		Farm size-yield relationship	
	ln(land available)	ln(total labor)	ln(output per ha)	
	(1)	(2)	(3)	(4)
Farmer productivity ($\ln s_i$)	0.199***	0.178***		
	(0.046)	(0.019)		
Farmer productivity ($\ln s_i$) × Western/Central region	0.288***	0.087***		
	(0.072)	(0.028)		
ln(land available)			-0.701***	-0.250***
			(0.023)	(0.025)
ln(land available) × Western/Central region			0.139***	0.131***
			(0.030)	(0.026)
Control for input ratio			No	Yes
No. obs.	2,269	15,264	10,255	10,059
R-squared	0.380	0.186	0.394	0.471

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%. All regressions use GPS measure of available land as proxy for farm size and include soil and farmer controls, as well as region-by-period and district fixed effects. Farm productivity s_i estimated using a flexible Cobb-Douglas in land and labor inputs with different parameters by region. Column 1 collapses the sample to one observation per farmer.

4.3 Farm size deeply confounded by market distortions

The broad literature on the inverse size-productivity relationship has had a profound influence on agricultural policy. To the extent that policy makers do not observe productivity (either land or farm productivity), but instead can easily observe farm size, the inverse size-productivity relationship provides a tractable mechanism for policy implementation.

We have shown that yields are not informative of the size-productivity relationship and hence can produce erroneous policy recommendations. But our evidence points to a more general conclusion: farm size is deeply confounded by distortions and hence, it is an inef-

However, these results do not rule that markets are imperfect nor that market distortions can play, at least partially, a role. They only imply that market distortions are not the *only* driver of this relationship.

fective instrument for policy. This conclusion is general because it applies to both measures of productivity. To illustrate this point, Table 7 documents the mean and dispersion of the two measures of productivity (farm productivity in logs and yields in levels) across farms within farm-size bins for different farm size categories. To characterize dispersion, in the case of farm productivity we use the difference between the 90th and 10th percentile, while for yields we use the ratio of the 90th and 10th percentiles.

The main observation is that there is substantial dispersion in both measures of productivity within a farm class, which is similar to or even greater than the dispersion of the overall distribution, and much greater than the dispersion in average productivity across farm sizes. For instance, within very small farms (0 to 1 ha), the dispersion in farm productivity between farms in the 90th and 10th percentiles is 2.362 (or a ratio of farm productivity of 10.6-fold), whereas the dispersion for the whole distribution of farm sizes is 2.108 (or a ratio of productivity of 8.2-fold). In comparison, the ratio of average productivity between the largest and smallest farm size classes is a factor of 2.25-fold. This implies that there is much more gain from reallocating resources across farms within a farm size class than across farms of different sizes. We observe a similar pattern when using the measure of yields (Y_i/T_i) for which the ratio of productivity between the smallest and largest size class is only a factor of 1.4-fold.

The implication of these results is that there is not a simple instrument for policy. Effective policy should facilitate better resource allocation by farm productivity, but productivity is difficult to observe for the policy maker. Our results suggest that policy should focus on fostering and improving markets, in particular, markets for land where even an egalitarian distribution of ownership rights can be decoupled from farm operational scales via rental markets or other more decentralized mechanisms. Decoupling land use from land rights can also have substantial effects on migration and occupation decisions, further contributing to productivity growth in agriculture (de Janvry et al. 2015, Adamopoulos et al. 2017).

Table 7: Productivity dispersion by farm size

Farm size (has)	% farms	Farm productivity (ln s_i) Mean	90th-10th percentile	Yields (Y/T) Mean	90th / 10th percentile
0-1	27.5	-0.289	2.362	3185.6	12.2
1-2	33.9	-0.030	2.010	2712.6	8.9
2-5	33.5	0.174	1.797	2386.0	6.5
5+	5.1	0.520	1.665	2274.0	8.1
All farms	100.0	-0.005	2.108	2698.5	8.8

Notes: Farm size classes are calculated using average area planted. Yields (Y/T) refer to average yields per farmer.

5 Conclusions

This paper does not attempt to explain the inverse yield-farm size relationship. Instead, we argue that using yields, as a proxy of productivity, is not informative of whether small farms are more or less productive. This happens because yields are affected by decreasing returns to scale and, importantly, by market distortions. These issues limit its usefulness to inform agricultural policies in developing countries, and may lead to wrong policy recommendations.

We illustrate these issues using the case of Uganda. We first show that the measure of productivity used can produce qualitatively different results. Yields suggest a negative farm size-productivity relationship. However, using estimates of farm productivity (from a production function), we find the opposite. These conflicting results do not seem to stem from omitted variables or measurement error. Instead, we show empirically that these differences arise because of the presence of DRS and market distortions.

References

Adamopoulos, Tasso and Diego Restuccia, "The size distribution of farms and international productivity differences," *American Economic Review*, 2014, *104* (6), 1667–97.

_ and _ , "Land Reform and Productivity: A Quantitative Analysis with Micro Data," Working Paper 25780, National Bureau of Economic Research April 2019.

_ , Loren Brandt, Jessica Leight, and Diego Restuccia, "Misallocation, selection and productivity: A quantitative analysis with panel data from China," Technical Report, National Bureau of Economic Research 2017.

Ali, Daniel Ayalew, Klaus Deininger et al., "Is There a Farm Size–Productivity Relationship in African Agriculture? Evidence from Rwanda," *Land Economics*, 2015, *91* (2), 317–343.

Assunção, Juliano J and Luis HB Braido, "Testing household-specific explanations for the inverse productivity relationship," *American journal of agricultural economics*, 2007, *89* (4), 980–990.

Assuncao, Juliano J and Maitreesh Ghatak, "Can unobserved heterogeneity in farmer ability explain the inverse relationship between farm size and productivity," *Economics Letters*, 2003, *80* (2), 189–194.

Bardhan, Pranab K, "Size, productivity, and returns to scale: An analysis of farm-level data in Indian agriculture," *Journal of political Economy*, 1973, *81* (6), 1370–1386.

Barrett, Christopher B, "On price risk and the inverse farm size-productivity relationship," *Journal of Development Economics*, 1996, *51* (2), 193–215.

_ , Marc F Bellemare, and Janet Y Hou, "Reconsidering conventional explanations of the inverse productivity–size relationship," *World Development*, 2010, *38* (1), 88–97.

Benjamin, Dwayne, "Can unobserved land quality explain the inverse productivity relationship?," *Journal of Development Economics*, 1995, *46* (1), 51–84.

Berry, R Albert, William R Cline et al., *Agrarian structure and productivity in developing countries: a study prepared for the International Labour Office within the framework of the World Employment Programme.*, Johns Hopkins Univ. Press, 1979.

Bevis, Leah EM and Christopher B Barrett, "Close to the edge: do behavioral explanations account for the inverse productivity relationship?," 2016.

Binswanger, Hans P, Klaus Deininger, and Gershon Feder, "Power, distortions, revolt and reform in agricultural land relations," *Handbook of development economics*, 1995, *3*, 2659–2772.

Carletto, Calogero, Sara Savastano, and Alberto Zezza, "Fact or artifact: The impact of measurement errors on the farm size–productivity relationship," *Journal of Development Economics*, 2013, *103*, 254–261.

Carter, Michael R, "Identification of the inverse relationship between farm size and productivity: an empirical analysis of peasant agricultural production," *Oxford Economic Papers*, 1984, *36* (1), 131–145.

Coldham, Simon, "Land reform and customary rights: the case of Uganda," *Journal of African Law*, 2000, *44* (1), 65–77.

Collier, Paul and Stefan Dercon, "African Agriculture in 50 Years: Smallholders in a Rapidly Changing World?," *World Development*, 2014, *63*, 92 – 101. Economic Transformation in Africa.

Desiere, Sam and Dean Jolliffe, "Land productivity and plot size: Is measurement error driving the inverse relationship?," *Journal of Development Economics*, 2018, *130*, 84–98.

Eswaran, Mukesh and Ashok Kotwal, "Access to capital and agrarian production organisation," *The Economic Journal*, 1986, *96* (382), 482–498.

Feder, Gershon, "The relation between farm size and farm productivity: The role of family labor, supervision and credit constraints," *Journal of development economics*, 1985, *18* (2-3), 297–313.

Funk, Chris, Pete Peterson, Martin Landsfeld, Diego Pedreros, James Verdin, Shraddhanand Shukla, Gregory Husak, James Rowland, Laura Harrison, Andrew Hoell, and Joel Michaelsen, "The climate hazards infrared precipitation with stations: a new environmental record for monitoring extremes," *Scientific Data*, 2015, *2.*

Gourlay, Sydney, Talip Kilic, and David Lobell, *Could the debate be over? errors in farmer-reported production and their implications for the inverse scale-productivity relationship in Uganda*, The World Bank, 2017.

Helfand, Steven, Matthew Taylor et al., "The Inverse Relationship between Farm Size and Productivity: Refocusing the Debate," Technical Report 2018.

Hopenhayn, Hugo A, "Entry, exit, and firm dynamics in long run equilibrium," *Econometrica: Journal of the Econometric Society*, 1992, pp. 1127–1150.

Hsieh, Chang-Tai and Peter J Klenow, "Misallocation and manufacturing TFP in China and India," *The Quarterly journal of economics*, 2009, *124* (4), 1403–1448.

Jr, Robert E Lucas, "On the size distribution of business firms," *The Bell Journal of Economics*, 1978, pp. 508–523.

Place, Frank and Keijiro Otsuka, "Land tenure systems and their impacts on agricultural investments and productivity in Uganda," *Journal of Development Studies*, 2002, *38* (6), 105–128.

Restuccia, Diego and Raul Santaeulalia-Llopis, "Land Misallocation and Productivity," Working Paper 23128, National Bureau of Economic Research February 2017.

_ , Dennis Tao Yang, and Xiaodong Zhu, "Agriculture and aggregate productivity: A quantitative cross-country analysis," *Journal of Monetary Economics*, 2008, *55* (2), 234–250.

Schlenker, Wolfram and Michael J. Roberts, "Nonlinear Effects of Weather on Corn Yields," *Review of Agricultural Economics*, 2006, *28* (3), 391–398.

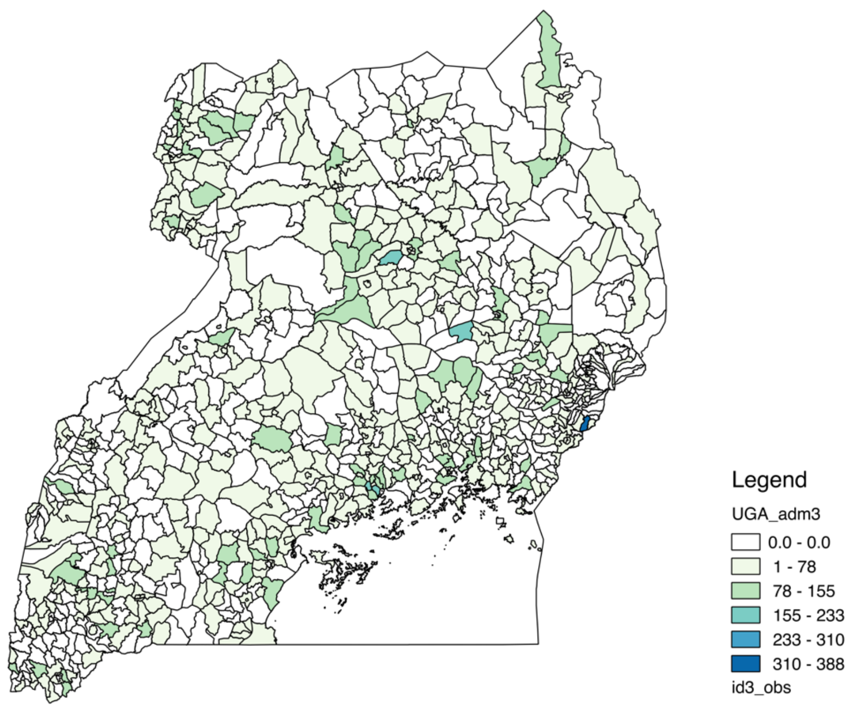
_ **and** _ , "Nonlinear temperature effects indicate severe damages to U.S. crop yields under climate change," *Proceedings of the National Academy of Sciences*, 2009, *106* (37), 15594–15598.

Sen, Amartya K, "An aspect of Indian agriculture," *Economic Weekly*, 1962, *14* (4-6), 243–246.

ONLINE APPENDIX - Not for publication

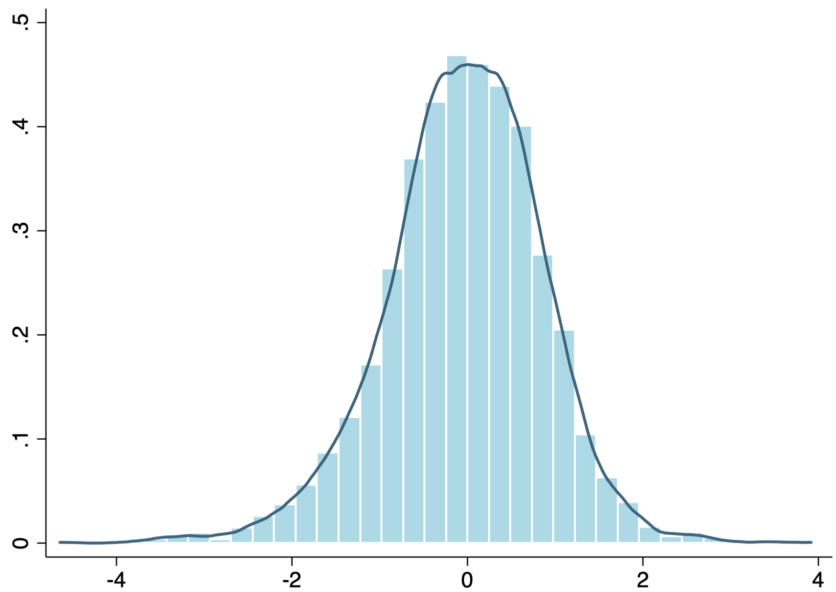
A Additional tables and figures

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 $\hat{\alpha} = 0.531$ and $\hat{\gamma} = 0.715$. The difference between the 90^{th} and 10^{th} percentile is 2.1.

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

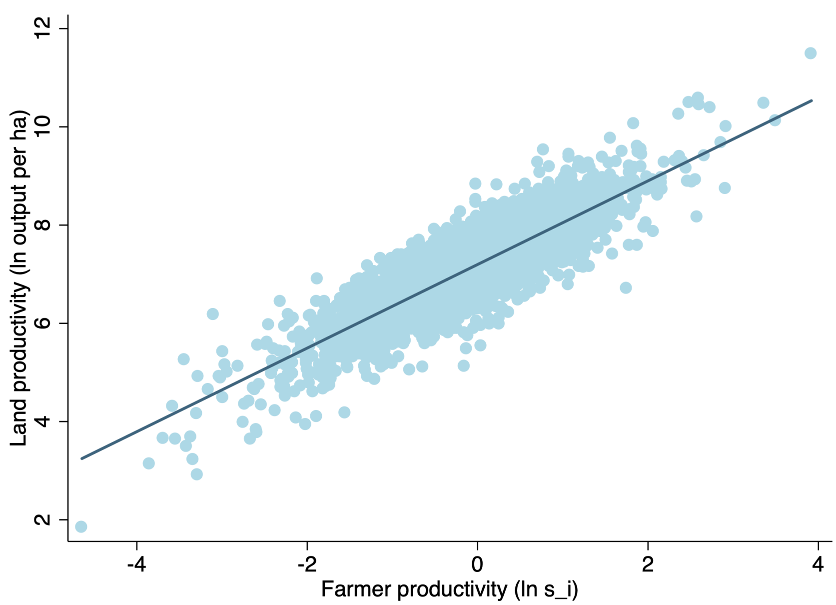
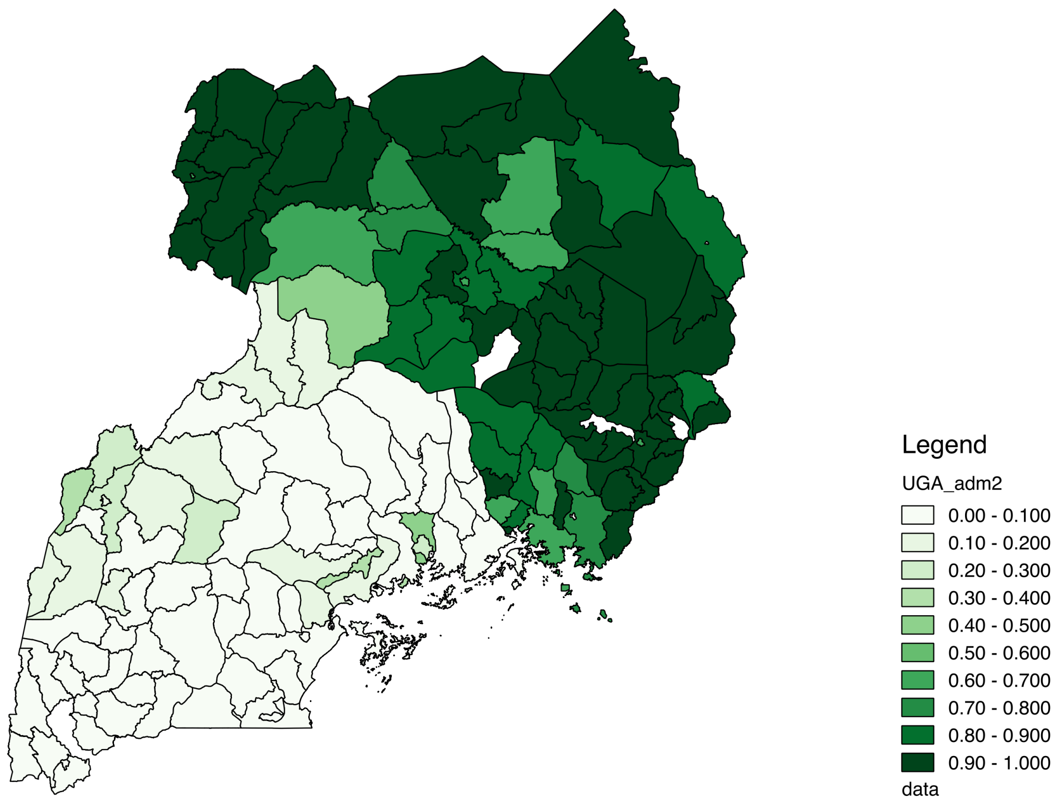


Figure A.4: Land tenure regimes in Uganda



Notes: Figure depicts the share of customary land in district (as % of agricultural land).

Table A.1: Production function estimates

	ln(output)		
	(1)	(2)	(3)
ln(land cultivated)	0.383***	0.392***	0.435***
	(0.020)	(0.020)	(0.068)
ln(total labor)	0.328***		
	(0.017)		
ln(domestic labor)		0.231***	0.239*
		(0.016)	(0.139)
ln(hired labor)		0.112***	0.110***
		(0.009)	(0.009)
Method	OLS	OLS	IV
Prod. function	CD	CD	CD + IV
Implied γ	0.711	0.735	0.783
Implied α	0.539	0.534	0.555
Observations	15,543	15,543	15,090
No. farmers	3,457	3,457	3,410

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 household and region-by-period fixed effects, plus weather controls, share of farm land using organic fertilizers, share of farm land using intercropping. CD = Cobb- Douglas. Column 3 uses land available and no. of household members who work in farm in last year as instruments for land cultivated and domestic labor. Land measured in has. Labor measured in person-days.

Table A.2: Using available land as measure of size

	ln(output/land cultivated)				Farm productivity	
	(1)	(2)	(3)	(4)	(5)	(6)
ln(land available)	-0.073***	-0.108***	-0.230***	-0.046**	0.167***	0.223***
	(0.013)	(0.013)	(0.022)	(0.021)	(0.011)	(0.020)
Controls	No	Yes	Yes	Yes	Yes	Yes
Household FE	No	No	Yes	No	No	No
No. obs.	16,010	14,530	15,739	3,252	16,370	3,249
R-squared	0.003	0.152	0.055	0.249	0.352	0.319

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 Table 2, as well as district fixed effects. Columns 1 to 4 also includes region-by-year 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.

Table A.3: Correcting for DRS and market distortions, using self-reported cultivated area

	ln(Y/T)		ln(Y/T) -($\hat{\gamma} - 1$)T	
	(1)	(2)	(3)	(4)
ln(T)	-0.257***	-0.064***	0.032**	0.225***
	(0.014)	(0.016)	(0.014)	(0.016)
ln(L/T)		0.390***		0.390***
		(0.017)		(0.017)
No. obs.	14,577	14,334	14,577	14,334
R-squared	0.175	0.217	0.152	0.195

Notes: Robust standard errors in parentheses. Standard errors are clustered at household level. * denotes significant at 10%, ** significant at 5% and *** significant at 1%. T = self-reported cultivated area. All regressions include district, region-by-year fixed effects and soil, weather and farmer controls as column 2 in Table 2. $\hat{\gamma} = 0.711$ obtained from Column 1 Table A.1.

Table A.4: Heterogeneity by % non-customary land in district

	Assessing misallocation		Farm size-yield relationship	
	ln(land available)	ln(total labor)	ln(output per ha)	
	(1)	(2)	(3)	(4)
Farmer productivity ($\ln s_i$)	0.171***	0.210***		
	(0.057)	(0.023)		
Farmer productivity ($\ln s_i$) × % non-customary land	0.343***	0.021		
	(0.102)	(0.044)		
ln(land available)			-0.676***	-0.211***
			(0.026)	(0.027)
ln(land available) × % non-customary land			0.110***	0.073*
			(0.042)	(0.038)
Control for input ratio			No	Yes
No. obs.	2,269	15,264	10,255	10,059
R-squared	0.379	0.186	0.393	0.469

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 use GPS measure of available land as proxy for farm size and include soil and farmer controls, as well as region-by-period and district fixed effects. Farmer productivity s_i estimated using a flexible Cobb-Douglas with different parameters by region. Columns 1 collapses sample to one observation per farmer. % non-customary land refers to the share of agricultural land in a district under a non-customary tenure regime.