Land productivity and plot size: Is measurement error driving the inverse relationship?

Paper prepared for the CSAE Conference 2017
Economic Development in Africa

Abstract: This paper revisits the decades-old puzzle of the inverse plot-size productivity relationship (IR), which states that land productivity decreases with increasing plot size in developing countries. While most empirical IR studies define yields as self-reported production divided by plot size, this paper complements this approach with an alternative, objective method to estimate yields: crop cuts. Using crop cuts as proxy for yields, the IR in Ethiopia disappears, while the negative relationship is strong when yields are based on self-reported production. The inverse relationship is even reversed as there exists a weak, positive correlation between plot size and crop cuts. This implies that self-reported production is systematically overestimated on small plots and underestimated on larger ones. Our findings suggest that the IR is an artifact of systematic measurement error in self-reported production. The implications of the rejection of an inverse *plot-size* productivity relationship for the inverse *farm-size* productivity relationship are discussed.

1. Introduction

With an estimated 40 percent of the world's extreme poor living in Sub-Saharan Africa (Ferreira et al., 2016),[1] the vast majority of whom derive their economic livelihood from agricultural activities (Livingston, et al, 2001), understanding land productivity is central to the global goals of poverty reduction and economic development. More directly demonstrating this point, Irz, Lin, Thirtle, and Wiggins (2001) estimate that a 10 percent increase in farmland productivity reduces the number of poor people in Africa by 7 percent. Despite the global importance of land productivity in determining the wellbeing of the poor, one very basic descriptive characteristic of land productivity remains a puzzle – that is why or even whether, there is an inverse relationship between farmland size and productivity?

Two empirical regularities between land size and land productivity have puzzled development economists for decades. The first regularity is well-known and is called the inverse *farm-size* productivity relationship (IR). As first noted by Chayanov (1926) in Russia and rediscovered by Sen (1962) in India, it states that land productivity decreases with farm size. This relationship has been observed in many developing countries and can be considered a stylized fact (Kimhi, 2006; Larson, Otsuka, Matsumoto, & Kilic, 2014). The inverse *farm-size* relationship has received considerable attention among researchers and policymakers because of its controversial implication for land reform (Collier & Dercon, 2013). If small farms are intrinsically more productive than larger farms, land redistribution would not only shift the distribution of wealth but also increase overall land productivity.

The second empirical regularity has received far less attention. We call it the inverse *plot-size* productivity relationship. This relationship, observed in many Sub-Saharan African countries, states that, within a household, smaller plots are more productive than larger plots (Ali & Deininger, 2015; Assunção & Braido, 2007; Barrett, Bellemare, & Hou, 2010). Although rarely emphasized, such a relationship implies that land fragmentation, that is subdividing the already small plots into even smaller units, would increase productivity. But more and smaller plots leads to increased travel time (in

[1] Estimates from PovcalNet (*iresearch.worldbank.org/PovcalNet*), the World Bank data tool for estimating extreme poverty, indicate that the majority of the poor in the rest of the world also live in rural areas. For example, essentially all of the extreme poor (living on less than USD 1.90 per day) in China are estimated to be in rural areas; and in India, this figure is estimated to be about 80 percent.

moving from plot to plot), more wasted space along borders, and reduced options for mechanization that typically functions best on larger plots. For these reasons and others, this finding seems non-credible. A common policy recommendation to improve agricultural productivity is for greater land consolidation (not fragmentation), and indeed land consolidation programs have been implemented in several countries (Ali & Deininger, 2015; Blarel, Hazell, Place, & Quiggin, 1992; Pašakarnis & Maliene, 2010).

The inverse *plot-size* and *farm-size* relationships are closely related. The most common and most intuitive explanation for the inverse *farm-size* relationship are missing markets. Missing land, labor, credit or insurance markets can explain differences in land productivity between households (Assuncao & Ghatak, 2003; Barrett, 1996; Carter & Wiebe, 1990; Eswaran & Kotwal, 1986; Feder, 1985). Households with little land, but an abundance of labor, apply too much labor to their land because missing land and credit markets prevent them from purchasing or leasing more land, while the labor markets are not sufficiently developed to rent out their labor. Indeed, several studies have shown that the inverse *farm-size* relationship disappears if labor input is valued at the market wage and have shown that land on small farms is more intensively cultivated than on larger farms (a point noted by many, including: Ali & Deininger, 2015; Benjamin & Brandt, 2002; Carter, 1984; Heltberg, 1998; Lamb, 2003). Missing markets cannot, however, explain differences in land productivity across plots within a household. In other words, missing markets can explain the inverse *farm-size* relationship, but fail to explain the inverse *plot-size* relationship. The existence of the inverse *plot-size* relationship has even led some to reject missing markets as the main explanation for the inverse *farm-size* relationship (Assunção & Braido, 2007).

Two other explanations are frequently suggested to explain both the inverse *plot-size* and *farm-size* relationship: soil quality and measurement error in self-reported plot size. Soil quality can explain the IR if smaller farms (plots) have on average more fertile soils than larger farms (plots) (Benjamin, 1995; Benjamin & Brandt, 2002; Bhalla & Roy, 1988; Chen, Huffman, & Rozelle, 2011). Though, in one of the few studies to use objective measures of soil quality data, Barrett et al. (2010) show that soil quality contributes only marginally to explaining the IR in Madagascar. They suggest that differences in labor inputs across plots or measurement error with respect to plot size are the only remaining candidate explanations for the inverse *plot-size* relationship. Lamb (2003) was the first to suggest that measurement error in self-reported land area could explain the inverse *farm* and *plot-size* relationship. Carletto, Gourlay and Winters (2015) find that replacing self-reported land size with GPS measurement weakened the *farm-size* relationship somewhat in several African countries, but the relationship never disappeared and even strengthened in Uganda (Carletto, Savastano, & Zezza, 2013).

This paper offers and tests a new explanation for the inverse *plot-size* productivity relationship: measurement error in self-reported production. All recent papers on the IR define yields as self-reported production divided by plot size (Ali & Deininger, 2015; Carletto, Gourlay, & Winters, 2015; Carletto et al., 2013; Larson et al., 2014).[2] One reason why systematic measurement error in self-reported production has not yet been considered as an explanation for the IR is because the data requirements to test it are very demanding. Systematic measurement error can only be studied if one has a second, independent measurement method of the same concept. In this study, we draw on two waves of a nationally representative dataset from Ethiopia which asked farmers to report production for all plots, and these plots were all measured with GPS (or with tape and compass for very small plots). In addition to the self-reported crop production, crop cuts were implemented on a limited set of randomly selected plots.

First used in the 1950s in India, the crop cut method is now widely recommended by the Food and Agricultural Organization of the United Nations for the accurate measurement of crop production (Fermont & Benson, 2011). The crop cut methodology estimates yields by first randomly sampling plots from a master listing of plots, and then from within each of the selected plots, randomly drawing a small subplot (often a 4m x 4m square), delimiting it, cutting, and weighing the harvest of this subplot. Yields are then defined as the harvest in the subplot divided by the area of the subplot.

While crop cuts are often times considered to be a gold standard for measuring crop production, that is not to suggest that they estimate yield without error. As one example, even though mechanical scales are much more precise than self-reports, they still entail rounding error when reading the weight from the scale. Or, as another example, because crop cutting entails randomly selecting a subplot from a larger plot, there is sampling error if production is heterogeneous within the plot. While we assume that these sorts of errors are mean zero, this assumption is not necessary for our findings to hold. The main assumption for our finding to hold is that whatever error there is in crop cut estimates, whether from sampling or measurement, the mean error is independent of the

[2] Most studies about the IR do not discuss how they measured yields, but all the studies cited above used self-reported production. As far as we know, no study has used crop cuts.

size of the plot. Given that the size of the subplot is fixed, regardless of the plot size (i.e. independent of the size of the plot), we maintain that whatever error there is associated with measuring the fixed-size subplot, it is similarly independent of the larger plot. Put another way, the mean error in crop cuts may be correlated with the size of the subplot, but there is no reason for why it should be related to the overall size of the plot from which the subplot is selected.[3]

The key finding in this paper is that when using crop cuts to measure output, the inverse productivity *plot-size* relationship disappears in our data. In contrast, and consistent with existing literature, when we use self-reports of crop production, the relationship is strongly negative. This finding suggests that production is systematically overreported on small plots and underreported on larger ones.

The remainder of this paper is structured as follows. In the next section, we discuss the data and provide some descriptive statistics. We then show the conditions under which measurement errors in self-reported production can generate the inverse *plot-size* productivity relationship and discuss the estimation strategy. In the result section, we show that the existence of the inverse *plot-size* productivity relationship depends on the method used to measure yields. Next, the sign and magnitude of measurement error in self-reported production as function of plot size is derived. Section 6 concludes and discusses the implications for the inverse *farm-size* relationship.

2. **Data**

We use data from two waves of the nationally representative Ethiopia Socioeconomic Survey (ESS).[4] This survey is an ongoing project to collect panel data in Ethiopia with both detailed information on household wellbeing and agricultural activities. It is implemented by the Central Statistical Agency of Ethiopia in close collaboration with the LSMS-ISA[5] team of the World Bank, which has a long history of producing high-quality household survey data. The first wave was administered in 2011/2012 to 3969 rural households, while the second wave was administered in 2013/2014 to 5262 household – 3776 panel households and 1486 new, mainly urban, households. We only included households in the dataset that were interviewed in both waves, so our sample is of rural

[3] In contrast, systematic correlation between measurement error in self-reported production (for the entire plot) and plot size cannot be ruled out (Carletto et al., 2015).
[4] All data and relevant documentation are publicly available at: go.worldbank.org/HWKE6FXHJ0
[5] LSMS-ISA: Living Standards Measurement Study – Integrated Surveys on Agriculture.

households. The time dimension in the data was not exploited, but pooling of the data both increases estimation power and allows us to show that the finding is not specific to a one point in time.[6] The survey collected standard information on household characteristics, consumption, living conditions and health.

The unique feature of the survey is its focus on agriculture. To gather detailed and accurate agricultural data at plot level, households were visited three times during the agricultural year. The first visit occurred in September-October to collect data on planting activities. During this visit, the area of most plots was measured with GPS.[7] The second visit occurred in November and implemented the livestock module. The final visit took place from January to April and collected data on agricultural production. This visit also included the household questionnaire.

In this paper, we exploit that yields were measured with two different methods: crop cuts and self-reported production by plot in combination with land measurement with GPS. In addition, we also used the detailed information on labor input at plot level during planting (first visit) and the harvest (third visit). Crop cuts were implemented in both waves of the survey for 23 major crops. Five plots per crop were randomly selected from a list of all plots cultivated by the sampled households within an enumeration area. In most cases, the plot selected for crop cutting was mono-cropped. Only if the crop was cultivated on fewer than five plots in pure stand within an enumeration area, were crop cuts also implemented on intercropped plots.

Once a plot was sampled from this list, a rectangular subplot within the plot was randomly sampled. In the first wave, this subplot was 2m by 2m square, while it was 4m by 4m square in the second wave.[8] Within the subplot, the crop was harvested by a trained enumerator and weighed. If logistical constraints allowed, crop cutting occurred simultaneously with the harvest of the main crop by the farmer. In both waves, nearly 40

[6] That is to say, data from two different time periods provides support to the view that the findings have "external validity" in a temporal sense. Also, given the changes in units used for self-reports across the waves (standardized units in wave 1, greater allowance for local units in wave 2), the use of both waves indicates that the finding is robust to these variations.
[7] The smallest plots were also measured with tape and compass, which is the gold standard of area measurement.
[8] Most enumerators participated in both waves and were intensively trained prior to the fieldwork. Yet, a researcher closely involved with the field work suggested that that some enumerators may have been confused and implemented crop cuts on a 2m x 2m square in wave 2, rather that the prescribed 4m x 4m. Such errors are unavoidable in large-scale surveys. As a consequence, yields may be underestimated in wave 2. Again though, this error was not associated with the size of the plot from which the subplot was sampled.

percent of the crop cuts were executed in November and more than 90 percent were executed between October and December. Both fresh weights and dry weights were recorded.[9] The correlation between both measures was over 0.95. We used the dry weights to calculate yields. 2975 and 3532 crop cuts were taken in wave 1 and 2, respectively. We discarded, however, those crop cuts for which we had fewer than 150 observations per crop. Finally, 5920 crop cuts remained in the dataset, providing yield estimates for 19 different crops.

Farmers reported the harvest per crop and plot during the third visit. In wave 1, most visits occurred in January (57%), while in wave 2 most visits took place in February (78%). Note that the recall period ranges up to 5 months since most crops during the Meher season are harvested from September till February (Taffesse, Dorosh, & Asrat, 2011). On average, production was reported 77 and 98 days after the implementation of the crop cuts in wave 1 and 2, respectively. Production was reported on 3683 and 23,638 plots in wave 1 and 2, respectively.

The design of the survey with respect to reporting the harvest differed substantially between the waves in three respects. First, in the first wave farmers reported the harvest in kilograms, while in the second wave farmers could opt to report in local units, which were converted into kilograms in a later stage. Second, in the first wave production data was only collected on plots where crop cuts were implemented, while production data was recorded for all plots in the second wave. This explains why we have significantly fewer observations in wave 1 than wave 2. Third, only in the second wave were farmers asked to estimate the share of land devoted to each crop if the plot was cultivated in mixed stand. As a result, yields of plots in mixed stand are more precisely estimated in wave 2 than wave 1. Yields in wave 1 for plots in mixed stand are by construction underestimated since we divided by total plot area. To take this into account, we always included an interaction term between the wave and the crop stand.[10] As a robustness check, the results are also reported by wave in order to show that the results are not driven by differences in survey design between the waves.

By gathering detailed data on labor input at the plot level, the ESS data also allows us to examine the hypothesis that differences in labor use at the plot level is the main

[9] For details on the methodology followed for the crop cut data used in this analysis, see http://go.worldbank.org/ZK2ZDZYDD0.

[10] Results do not change if the sample is restricted to mono-cropped plots. Results available upon request.

source of the IR (Barrett et al., 2010). The ESS included a labor module on planting activities (first visit) and harvest activities (third visit). Farmers were asked to report family labor, hired labor and exchange labor by plot, and this information is included in the regression analysis to address the hypothesis that this is the reason for the IR. Furthermore, differences in labor input may also explain differences in self-reported land productivity and crop cuts, a point that will be elaborated in the methodology section.

Given our focus on the inverse *plot*-size productivity relationship, the unit of analyses is the plot. In the ESS, enumerators first defined parcels, which are units of land that are owned by a single household and surrounded by land owned by another household or demarcated by natural boundaries (forest, water, road). Within a parcel, plots were then identified. In most cases, plots were clearly demarcated by hedges or paths, and most plots (80%) were single-cropped. All data (land area, production, inputs and crop cuts) were collected at plot level, with the exception of irrigation, soil quality and the possession of a land certificate, which were reported at parcel level. Since parcels are already small pieces of land, plots are even smaller. Mean and median plot size is 1300m^2 and 640m^2, respectively and more than 95 percent of the plots were smaller than 5000m^2 (0.5ha).

3. Descriptive statistics

In order to examine if crop cuts were indeed randomly implemented across households and plots, we compare households with at least one plot selected for crop cutting versus those without a single plot selected for crop cutting and compare plots with and without crop cuts.[11] The validity of the main finding in this paper does not rest on the random allocation of the crop cuts across households since we also compare differences within the sample of crop cut households, but the random allocation is what allows us to extend our inferences to the rural population of Ethiopia.

Household characteristics are very similar between both groups (table 1). The only important difference is that households selected for crop cuts owned slightly more land (1.30 ha versus 1.19 ha). This is in line with expectations though, since households with more land are also more likely to cultivate at least one plot suited for crop cutting.

[11] Sampling weights are not used in the descriptive statistics nor in the regressions, although the standard errors are corrected for clustering at the enumeration area. The findings are robust to using sampling weights from wave 1 or wave 2. Results are available upon request.

Table 1: Household characteristics for households with and without plots with crop cuts

	No crop cuts	Crop cuts	Significant difference (t-value)	
Landholdings (ha)	1.18	1.34	3.07	***
Applied chemical fertilizer (%)	47	50	1.29	
Asset index	0.16	0.17	2.00	**
Household size	5.75	5.90	1.62	
Age household head	46.47	45.67	1.42	
Household head can read and write (%)	38	39	0.73	
Female headed household (%)	20	17	1.50	
N (min/max)	2954/3018	878/890		

Notes: Number of observations differs by variable due to missing variables. T-values estimated by clustering the standard errors at the enumeration area. ***, **, * denote statistical significance at the 1%, 5% and 10% levels.

In terms of plot characteristics, differences between plots selected for crop cuts and the other plots are also potentially due to the sampling design (table 2). Mono-cropped plots are more likely to be further away from the dwelling of the household and therefore crop cuts tended to be somewhat further away as well. Differences in terms of fertilizer application (both organic and inorganic), irrigation, labor input and soil quality are small. The only exception is family labor during the harvest. Plots without crop cuts appear to be more intensively harvested (12.2 days/ha) than plots with crop cuts (8.1 days/ha).

In order to explore if an inverse productivity plot-size relationship exists in our data, we examined maize yields in wave 2 by quartile of plot size and measurement method (figure 1). Mean and median yields based on self-reported production clearly decrease with increasing plot size. Median maize yields are over 2000 kg/ha on the smallest plots, but decrease to 1000 kg/ha on the largest plots. This pattern holds more generally in the data. Using all observations, a simple, bivariate regression confirms the existence of the inverse plot-size productivity relationship. Regressing the log of yields[12] on the log of plot size shows that yields decrease by 34 percent if plot size doubles (t-value=73).

[12] Production was converted in its monetary value using local prices. If we had at least 10 observations, median self-reported prices by crop were calculated at the Woreda level, which is the administrative unit just above the lowest unit (the Kebele). If there were fewer than 10 observations, the median national price was used.

Table 2: Plot characteristics for plots selected and not selected for crop cutting

	No crop cut	Crop cut	Significant difference (t-value)	
Plot size (m²)	1227.30	1501.00	6.60	***
Distance from household dwelling (km)	0.68	0.95	5.96	***
Plot slope (%)	14.55	14.35	0.45	
Plot elevation	1931.35	1968.14	2.10	**
Plot potential wetness index	12.58	12.55	0.52	
Land title (certificate) (%)	48	50	1.07	
Pure stand (%)	57	87	18.26	***
Applied manure (%)	33	21	8.25	***
Applied compost (%)	5	6	1.05	
Applied organic fertilizer (%)	2	2	0.36	
Irrigation (% of plots)	4	2	3.04	***
Self-reported soil quality (only wave 2)	1.75	1.89	6.45	***
Fertilizer (kg/ha)	44.40	46.21	0.54	
Family labor planting (days/ha)	16.89	16.13	0.99	
Hired labor planting (days/ha)	7.48	7.27	0.18	
Exchange labor planting (days/ha)	15.19	15.08	0.09	
Family labor harvesting (days/ha)	12.23	8.14	7.95	***
Hired labor activities (days/ha)	4.81	4.16	1.11	
Exchange labor activities (days/ha)	15.33	15.08	0.14	
N (min/max)	20046/21401	3296/5920		

Notes: Number of observations differs by variable due to missing variables. Soil quality was only reported in wave 2. T-values estimated by clustering the standard errors at the enumeration area. ***, **, * denote statistical significance at the 1%, 5% and 10% levels.

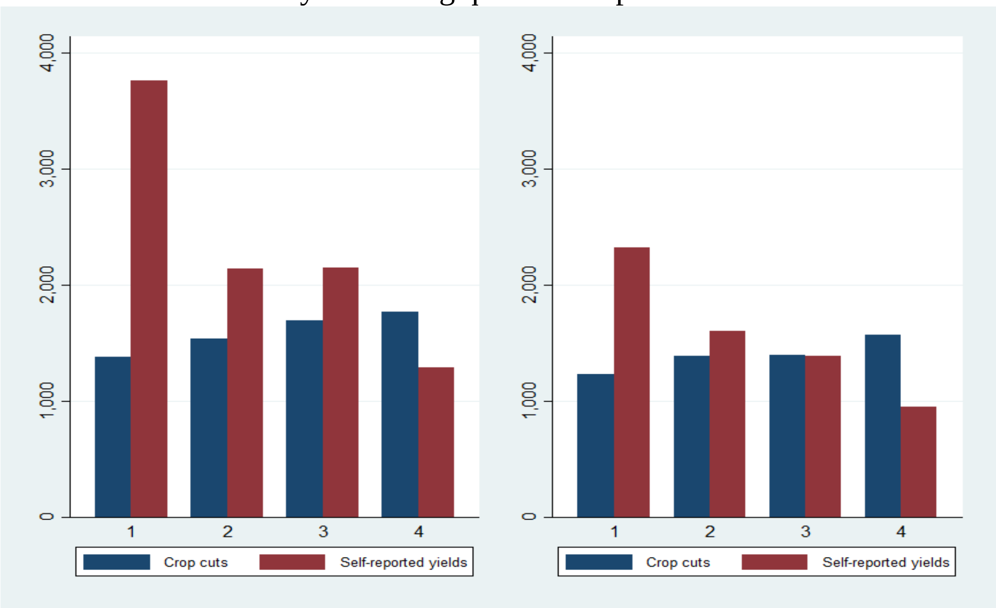
The inverse productivity plot size relationship disappears, however, when yields are estimated with crop cuts. Crop cuts of maize even tend to increase with plot size. Median maize yields equal 1200 kg/ha on the smallest plots and increase to 1800 kg/ha on the larger plots. This pattern is again observed for all crops. The bivariate regression suggests that crop cuts increase by 9 percent if plot size doubles (t-value=10).

When comparing crop cuts with yields based on self-reporting, we note that yields based on crop cuts are substantially lower than self-reported yields on the smallest plots, but higher on the largest plots. This suggests that self-reported production is systematically overreported on small plots and underreported on larger ones.[13] Given the few important differences between households and plots with or without crop cuts, this suggests that the inverse relationship can be attributed to systematic measurement error

[13] It is interesting to note that this same pattern of systematic divergence between self-reports and objective measures also occurs when comparing GPS measures of plot size with self-reports (Carletto et al., 2013).

in self-reported production. This will be examined in more detail in the remainder of this paper.

Figure 1: Mean (left panel) and median (right panel) maize yields (kg/ha) in wave 2 by measurement method by increasing quartiles of plot size



Notes: Sample restricted to plots with both crop cuts and self-reported production. 62, 134, 138 and 134 observations in quartile 1 to 4, respectively. The quartiles have an average plot size of 150m², 425m², 1128m² and 3645m², respectively.

4. Methodology

We develop a simple model to explain how systematic measurement error in self-reported production can generate an inverse plot-size productivity relationship. Traditionally, the IR is estimated by regressing yields on plot size, A, and a set of plot characteristics, X_i:

$$\log(yield_i) = \alpha \log(A_i) + \beta X_i + v_i \tag{1}$$

where i represents the plot, and v_i is a normally distributed error term. The parameter of interest is α, which reflects the correlation between plot size and yields. An inverse relationship can be rejected if $\alpha \geq 0$.[14]

The dependent variable in the equation are 'observed' yields, which are usually estimated as self-reported production, X_i, divided by plot size, A_i. Self-reported production equals the sum of 'true', but unobserved production, X_i^*, and an error term, ε_i.

[14] α is nearly always assumed to be independent of plot size. If α depends on A, equation 1 is non-linear, which complicates the estimation procedure (see (Verschelde, D'Haese, Rayp, & Vandamme, 2013)) for an application).

Hence, the following identity links observed yields ($yield_i$) with the true yields ($yield_i^*$) and measurement error in self-reported production:

$$yield_i = \frac{X_i}{A_i} = \frac{X_i^* + \varepsilon_i}{A_i} = yield_i^* \ + \frac{\varepsilon_i}{A_i} \qquad (2)$$

While the measurement error in self-reported production ε_i is plot specific, true yields, $yield^*$, are assumed to be constant conditional on inputs such as fertilizer and plot characteristics such as soil quality.

Only certain types of measurement error generate an inverse productivity plot-size relationship. This can be illustrated as follows. Assume that the inverse relationship does not exist. This implies that $corr(yield_i^*, A) = 0$. Consequently, a spurious inverse relationship will be observed if and only if:

$$corr(yield_i, A) < 0 \quad \Leftrightarrow \quad corr(yield_i^* + \frac{\varepsilon_i}{A}, A) = corr\left(\frac{\varepsilon_i}{A}, A\right) < 0 \quad (3)$$

Equation 3 summarizes the condition that the error term in self-reported production needs to meet in order to generate an inverse relationship. First, it shows that "classical" measurement error (i.e. mean zero and independent of A) will not result in an inverse relationship. Second, in order to generate the IR, relative measurement error (i.e. the ratio of the error term to true production, or $\frac{\varepsilon}{A\, yield^*}$), needs to decrease with plot size.

For heuristic purposes, consider measurement error which is a non-stochastic, positive scalar, (ε). In this case, farmers overreport harvest on all plots by the same amount (ε), regardless of plot size. As such, equation 3 simplifies to $\varepsilon * corr\left(\frac{1}{A}, A\right)$ <0.[15] If measurement error is constant, the inverse relationship is generated because a small, positive measurement error in self-reported production biases estimated yields more on a small plot than a larger one. An example clarifies this process. Assume that farmers overestimate production by 10kg on all plots and that 'true' yields are 1000kg/ha. True production on a plot of 100m² is then 10kg, but farmers report a production of 20kg. Observed yields thus equal 2000 kg/ha, which overestimates true yields by 100 percent. On a plot of 1000m² true production is 100kg, but a harvest of 110kg is reported. On this plot, observed yields are 1100 kg/ha and overestimate true

[15] The correlation $corr(A, 1/A)$ is always negative because of Jensen's inequality, which states that $\varphi\big(E(X)\big) \leq E(\varphi(X))$, where φ is a convex function. Hence, $corr\left(A, \frac{1}{A}\right) = 1 - E(A)E(\frac{1}{A}) \leq 0$ since $\varphi\big(E(A)\big) \leq E\big(\varphi(A)\big)$, where φ is the convex function that transforms A to $1/A$.

yields by 10 percent. Consequently, while true yields are the same on the small and large plots, observed yields are 2000 kg/ha on the small plot and 1100 kg/ha on the larger plot and an inverse plot-size productivity relationship is observed.

Other types of measurement error also satisfy the condition expressed in equation 3 and, therefore, generate a spurious relation between plot size and productivity. For instance, farmers could systematically overreport production on small plots and underreport it on larger ones. Underreporting on larger plots may occur if farmers fear taxation or do not want to reveal their wealth.[16]

While measurement errors in self-reported production could induce a spurious inverse relation, we argue that measurement errors in crop cuts are unlikely to do so. Crop cuts follow a standardized, objective design, which is independent of plot size. The methods followed for a crop cut from a sub-plot are exactly the same whether the plot is large or small. We therefore assume that whatever measurement error exists in crop cuts is independent of plot size.[17] Hence, if measurement error in self-reported crop cuts explain the IR, this relationship should disappear once crop cuts are used as proxy for yields.

One caveat deserves some attention. It is possible that crop cuts and yields based on self-reporting do not capture the same underlying concept. In other words, the true, unobserved yield, represented by $yield^*$ in equation 2, is not necessarily the same for both methods. Crop cuts measure the maximum potential yield, while self-reported production measures the actual harvest. In theory, crop cuts are an upper bound of yields, while self-reported production gives a more accurate estimate of the realized yields since it takes harvest losses into account. This conceptual difference is an important challenge to our bi-variate regression analysis. If the IR holds for self-reported proxies of land productivity, but disappears for crop cuts, one may argue that smaller plots are simply

[16] One intuitive type of measurement error that will not generate a spurious inverse relationship is when the measurement error is proportional to true production. In this case, the error equals $\varepsilon_i = \delta A_i yield^*$. Using equation 2, this implies that observed yields are always equal to $(1 + \delta) yield^*$ and are constant across plot size.

[17] As further example of measurement error that might be attached to crop cuts, the time that one allows a crop to dry affects the measured output of the crop cut. If this time varies across crop cuts, then this introduces measurement error. Indeed, this is the motivation for why the data records whether the crop was measured wet or dry, and is why we run separate regressions. A somewhat related example is based on the fact that crop cuts are for a random subplot of a larger plot. Given that yields will vary across sub-plots within a plot due to a variety of reasons, this random selection of the sub-plot introduces sampling error. But, the expectation of these sampling errors is by definition mean zero and is not correlated with plot size.

more completely harvested. To control for this possibility, we include labor input during planting and harvest at plot level in the regressions. In addition, we estimate the correlation between plot size and labor input in order to examine if small plots are indeed more completely harvested than larger ones.

The previous discussion about biased measurement of yields showed that the inverse relationship can only be generated by measurement error if relative measurement error in production decreases with plot size. The sign and magnitude of the relative measurement error can be estimated. To do so, the log of the ratio (R) of yields based on self-reporting and crop cuts are regressed on plot size:

$$\log(\mathrm{R}) = \log\left(yield_i^{SR}/yield_i^{CC}\right) = C + \delta \log(A_i) + \beta L_i + v_i \tag{4}$$

Labor input is included in the regression to capture the difference in the definition of the underlying 'true' yields. Using this estimation, we can predict the ratio, $R(A)$, and use it to estimate relative measurement error as a function of plot size. Manipulating equation 2 gives the following relation between relative measurement error and plot size (A):

$$\frac{\varepsilon}{A_i * yield^*} = vR(A) - 1 \tag{5}$$

Where v is the ratio of yields based on crop cuts to true yields. Previous research has shown that crop cuts are likely to overestimate 'true' yields, and we thus believe that $v > 1$ (Fermont & Benson, 2011). Since R is decreasing in plot size, the relative measurement error will also decrease with plot size. Moreover, at a certain threshold of plot size (i.e. if $vR(A) = 1$) measurement error is negative implying that observed production is lower than true, unobserved production. This critical threshold depends on assumption about v as well as on the parameter δ.

The estimation of the inverse productivity plot-size relationship (equation 1) typically includes household fixed effects (Assunção & Braido, 2007). In other words, one examines if small plots are more productive than large plots within the same household. We cannot include household fixed effects when using crop cuts as dependent variable since only on rare occasions were several crop cuts implemented on plots from the same household. We do, however, include enumeration area fixed effects[18] as well as a set of household characteristics.

[18] Interestingly, enumeration area fixed effects also control for enumerator bias. The survey was organized in such a way that there was a different enumerator for every enumerator area.

5. Results

Table 3 shows the results of the estimation of equation 1 using yields based on self-reported production (columns 1 - 4) and yields based on crop cuts as the dependent variable (columns 5 - 6). In all specifications, robust t-statistics are reported to account for heteroscedasticity. We first discuss the results for self-reported yields and establish that the inverse relationship holds in the data. Next, we show that the inverse productivity plot-size relationship disappears when yields are measured with crop cuts. We then examine if differences in labor input between small and large plots or systematic measurement error in self-reported production explain the findings.

The classic approach to estimate the inverse productivity plot-size relationship includes household fixed effects and plot characteristics (table 3, column 1). This shows that yields decrease by 40 percent if plot size doubles – indicating that the inverse relationship holds in Ethiopia and is stronger than in many other Sub-Saharan African countries (Larson et al., 2014). The second specification replaces household fixed effects with enumeration area fixed effects and includes controls for household characteristics. This does not affect the IR. In other words, the results remain similar when examining differences in land productivity between large and small plots within a household or between households within a same enumeration area. This is important because household fixed effects are not included when using crop cuts as proxy for land productivity. Restricting the sample to plots for which we have both self-reported production and crop cuts (columns 3 and 4), reduces the sample size from over 25,000 observations to slightly over 5000 observations. In both specifications, the IR remains highly significant. Specification 4 includes as an additional control labor at plot level, and this does substantially weaken the magnitude of the IR, but the negative relationship continues. Without including labor, yields decrease by 30 percent if plot size doubles, while yields only decrease by 16 percent if differences in labor input between small and large plots are taken into account. This important point will be discussed in more detail below.

In addition to plot size and labor, there are several plot and household characteristics that are significantly correlated with yields based on self-reported production (see full results in appendix). Of the 11 plot characteristics included in the regressions, 4 are significantly correlated with yields. Irrigation and the application of fertilizer as well as the distance to the dwelling increase yields. In wave 2, yields are higher on plots with mixed cropping systems, perhaps because of complementarities between crops. In wave 1, yields are by design always underestimated on mixed plots, which is confirmed in the regressions. With regards to household characteristics, only the asset index is significantly correlated with yields in most specifications. The asset index is a proxy for household's wealth.[19] Perhaps unsurprisingly, wealthier households tend to have higher yields. In some specifications, female headed households have lower yields, while poorly educated household have higher yields.

The results of the estimation of the IR using crop cuts as the dependent variable are also shown in table 3 (columns 5 - 6). Specification 5 is the counterpart of specification 3, while specification 6 includes labor input and can be compared directly to specification 4. In both specifications the IR disappears. We even observe a positive and highly significant correlation between plot size and yields. Several other variables are also correlated with crop cuts. As in the previous specifications, chemical fertilizers and the asset index increase yields, although the latter variable is only marginally significant. Labor input also correlates positively with crop cuts, but the correlations are weaker than between labor and yields based on self-reporting.

Table 3: The inverse plot-size relationship for yields, self-reports and crop cuts

	Self-reported measurement				**Crop cuts**	
	(1)	(2)	(3)	(4)	(5)	(6)
Log plot size (m²)	-0.397***	-0.396***	-0.303***	-0.161***	0.104***	0.149***
	(-33.26)	(-36.11)	(-12.65)	(-6.22)	(8.63)	(9.36)
Household fixed effects	X					
Enumeration area fixed effects		X	X	X	X	X
Household characteristics		X	X	X	X	X
Labor input				X		X
Observations	25811	25004	5248	5059	5248	5059
R-squared	0.194	0.201	0.110	0.215	0.120	0.140

Notes: ***, **, * denote statistical significance at the 1%, 5% and 10% levels. T-statistics in parentheses. Errors clustered at the enumeration area. Full results are provided in the appendix. All regressions include a dummy for the wave and plot characteristics.

Several robustness checks indicate that our findings remain consistent across different subsamples. Table 4 reports the results of three important robustness checks. All regressions include household and plot characteristics as well as labor input at plot level. First, we estimated the IR separately for wave 1 and wave 2 of the survey. This is important because the methodology changed between the waves (and it also allows us to suggest that the findings are not driven by some temporally-specific factor). This is

[19] The asset index is a simple, standardized count of the number of assets owned by the household. Thirty-five assets are considered including agricultural equipment such as ploughs, sickles and axes as well as durable consumption goods such as radios and mobile phones.

particularly a concern for yields based on crop cuts since a crop cut was implemented on a 2m x 2m square in wave 1, while a subplot of 4m x 4m was sampled in the second wave. The change in methodology may explain the sharp decrease (by 30%) in average crop cuts between wave 1 and 2, while self-reported production per hectare only decreased by 10 percent between the waves.[20] In both waves, we observe a negative correlation between plot size and self-reported productivity and a positive correlation between plot size and crop cuts. As a second robustness check, we limited the sample to plots cultivated with maize. Rather than expressing maize yields in their monetary value, we expressed them in kg/ha. This allows us to examine if the results are influenced by measurement error in price levels. Again, the IR is strong for yields based on self-reporting, but disappears for crop cuts. In contrast to the previous findings, the positive correlation between plot size and crop cuts of maize is no longer significant at conventional levels. As a third robustness check, we assessed the impact of outliers. We followed the strategy of Larson et al. (2014) and excluded the bottom and top 5 percent of yields. This weakens the IR based on self-reported yields markedly, but has only a limited effect on the correlation between plot size and crop cuts. Outliers appear to be more of an issue for self-reported production than for crop cuts.

Table 4: Estimation of the IR for several subsamples

	Wave 1		**Wave 2**		**Maize**		**Discarding outliers**	
	Self-reported	Crop cuts	Self-reported	Crop cuts	Self-reported	Crop cuts	Self-reported	Crop cuts
Log plot size (m²)	-0.160***	0.171***	-0.129***	0.138***	-0.439***	0.060	-0.100***	0.110***
	(-3.67)	(7.78)	(-4.33)	(7.34)	(-7.31)	(1.50)	(-4.91)	(8.27)
Observations	1860	1860	3152	3152	726	722	4664	4557
R-squared	0.198	0.101	0.249	0.094	0.356	0.182	0.168	0.133

Notes: ***, **, * denote statistical significance at the 1%, 5% and 10% levels. T-statistics in parentheses. Errors clustered at the enumeration area. Full results are reported in the appendix. All regressions include a dummy for the wave, household and plot characteristics, labor input and enumeration area fixed effects. Dependent variables are expressed in Birr/m², except for maize yields which are expressed in kg/ha. Samples are always restricted to plots for which we have both crop cuts and self-reported production.

A first explanation for the finding that the existence of the IR depends on the method used to estimate yields is that the different methods do not capture the same underlying concept. Crop cuts measure maximum attainable yields, while self-reported

[20] A potential explanation is 'edge effects'. Plants that grow on the border delimiting the subplot might be systematically harvested. Since this border is larger for small than larger plots proportionally to the size of the plot, edge effects are more substantial on smaller plots. This suggests that the measurement error in crop cuts may be correlated with the subplot size (but again, there is no reason for it to be correlated with the overall plot size).

production is a proxy for the actual harvest. The observed inverse relation between plot size and self-reported yields is a real phenomenon if small plots are more completely harvested than large ones. This would not rule out that potential yields, as measured by crop cuts, remain constant across plots. This hypothesis is only partially upheld in the data.

Table 5 shows the results of estimating the relation between labor input per hectare and plot size. Two types of labor (family and hired labor) and two agricultural activities (planting and harvesting) are distinguished. There is a strong, inverse relation between family labor (during planting and harvest) and plot size, and a positive, but less robust, relation between plot size and hired labor. This confirms that small plots are more intensively cultivated than large plots. Both findings are not novel in the literature and have been reported by Ali and Deininger (2015) in Rwanda and Larson et al. (2014) in several African countries.

Table 5: Estimation of the inverse relation between labor input per hectare and plot size

	(1) Family labor planting	(2) Hired labor planting	(3) Family labor harvest	(4) Hired labor harvest
Log plot size (m²)	-0.382***	0.0661***	-0.311***	0.0769***
	(-40.18)	(8.30)	(-22.08)	(8.01)
Observations	24743	24743	25004	25004
R-squared	0.299	0.038	0.169	0.035

Notes: ***, **, * denote statistical significance at the 1%, 5% and 10% levels. T-statistics in parentheses. Errors clustered at the enumeration area. Full results are reported in the appendix. All regressions include a dummy for the wave, household and plot characteristics, labor input and enumeration area fixed effects. All dependent variables are expressed in days/m².

While controlling for labor input during planting and harvest substantially weakens the IR, the relationship remains statistically significant (table 3, column 4). Family labor, particularly during the harvest, is correlated with yields based on self-reporting and crop cuts, but the former correlation is much stronger than the latter. This confirms that self-reported productivity is higher than crop cuts if the plot is more completely harvested.[21] In sum, the analyses suggest that households indeed cultivate and harvest small plots more intensively than large plots, which contributes to the inverse relationship, but is not sufficient to explain the relationship.

[21] There is a clear endogeneity problem. Yields are higher because the plots have been more intensively harvested, but plots are also more intensively harvested if the harvest was good. We do not aim to establish causal relations since we only attempt to show that the existence of the IR depends on the method used to measure yields.

The second explanation for the IR is systematic measurement error in self-reported production. This can explain the IR if relative measurement error in self-reported production decreases with plot size. Regressing the ratio of self-reported yields versus crop cuts on plot size (table 6) confirms that yields based on self-reporting decrease with plot size relative to yields based on crop cuts.

Table 6: Log of the ratio of self-reported yields versus crop cuts regressed on plot size

	(1) Full sample	(2) Full sample	(3) Core sample
Log plot size (m^2)	-0.295***	-0.287***	-0.233***
	(-25.64)	(-11.99)	(-10.36)
Wave	0.414***	0.319***	0.291***
	(15.58)	(5.58)	(5.48)
Labor inputs (logs)			
Family labor planting (days/ha)		0.0302*	0.0305*
		(1.69)	(1.83)
Hired labor planting (birr/ha)		0.0119	0.00698
		(0.93)	(0.60)
Exchange labor planting (days/ha)		-0.00517	-0.00490
		(-0.54)	(-0.50)
Family labor harvest (days/ha)		0.226***	0.182***
		(9.70)	(7.90)
Hired labor harvest (birr/ha)		0.0654***	0.0461***
		(4.46)	(2.80)
Exchange labor (days/ha)		0.0644***	0.0506***
		(6.38)	(5.92)
Enumeration area fixed effects		X	X
Household characteristics		X	X
Plot characteristics		X	X
Observations	5914	5053	4593
R-squared	0.133	0.223	0.161

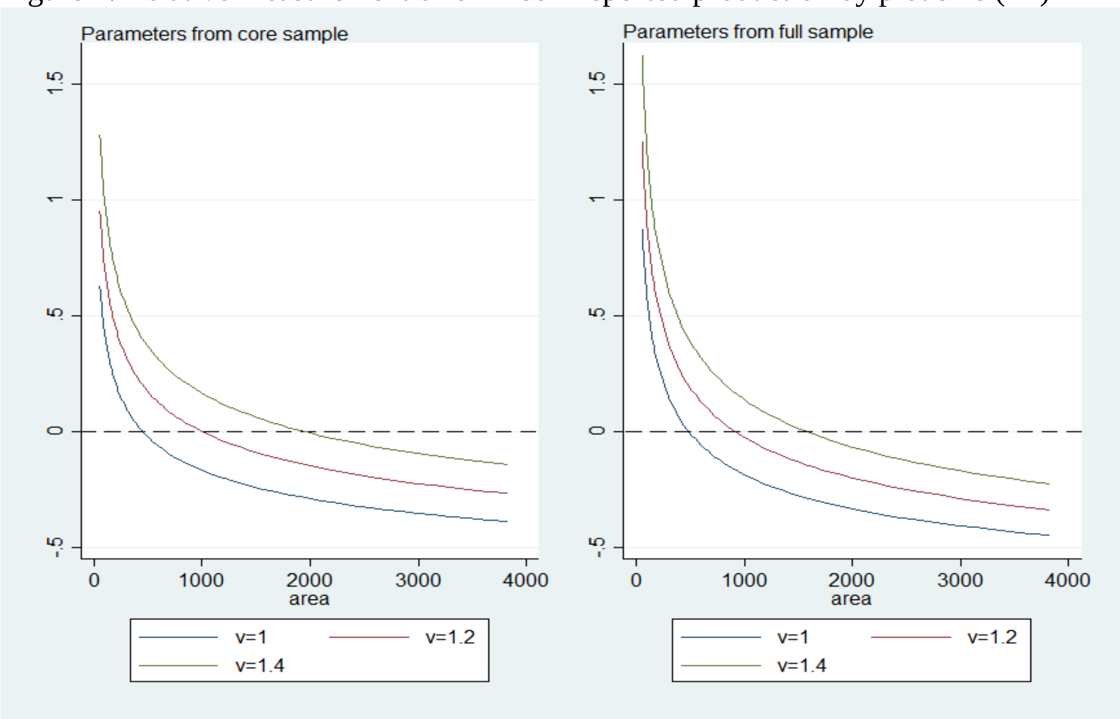
Notes: ***, **, * denote statistical significance at the 1%, 5% and 10% levels. T-statistics in parentheses. Errors clustered at the enumeration area. The core sample discard bottom and top 5% of yields based on self-reporting and yields based on crop cuts.

Based on these results and using equation 4 from the previous section, relative measurement error in self-reported production can be estimated as a function of plot size. This requires an assumption about the parameter v, which represents the ratio of yields based on crops cuts versus unobservable true yields. We show results for three values of v $(1;\ 1.2;\ 1.4)$. This choice corresponds to assuming that crop cuts are the gold standard of yield measurement (v=1) or systematically over estimate yields by 20 percent ($v = 1.2$) or 40 percent ($v = 1.4$). In order to assess the sensitivity of the results to the estimated parameters of the IR, relative measurement error in self-reported production is estimated using parameters for the core sample (left panel in figure 2), which excluded the top and bottom 5 percent of yields, and for the full sample (right panel in figure 2).

Independent of the choice of v, the graphs show that relative error in self-reported production decreases rapidly with plot size. Relative errors are smaller when parameter estimates are based on the core sample compared to estimates based on the full sample, but the trends are broadly similar. We will therefore only discuss the results for the core sample.

With $v = 1$, production is overreported by at least 50 percent on plots smaller than 80m² and underreported by more than 25 percent on plots larger than 1500m² relative to the 'true' production. For larger values of v, overreporting on small plots increases, while underreporting on larger plots decreases. The critical threshold of plot size below which overreporting of production changes to underreporting of production depends on the choice of v. This critical threshold equals 445m², 992m² and 2000m² for $v = 1, 1.2$ and 1.4, respectively.

Figure 2. Relative measurement error in self-reported production by plot size (m²)



Notes: The vertical axis represents relative overreporting or underreporting of production by the respondent versus crop cuts. The dotted, horizontal line demarcates overreporting from underreporting. Blue, red and green curves represent different assumptions with respect to the ratio of 'true' yields to 'crop cuts'.

With an additional assumption about 'true' yields, absolute measurement errors in self-reported production can be estimated. Assume that average 'true' yields are 1000 kg/ha. Under this assumption, true production on a plot of 100m² is 10kg and overreporting by 50 percent implies an overestimation by 5kg. With the same assumptions, true production on a plot of 2500m² is 250kg, but is underreported by 80kg. While an overestimation by 5kg on small plots seems reasonable, the large underestimation on larger plots appears less credible. It is worth pointing out, however, that a similar pattern has been observed for self-reported land area: farmers tend to overestimate the area of small plots, but underestimate the area of larger plots (Carletto et al., 2015). Moreover, the order of magnitude of over and underestimation is similar to our findings. For instance, Carletto et al. (2015) report that farmers overestimate plot size by 103 percent relative to GPS measurement on plots smaller than 0.5 acres (2000 m²) and underestimate it by 33 percent on plots larger than 5 acres (20,000 m²).

6. Conclusion

This paper revisits the inverse *plot-size* productivity relationship and presents evidence to reject it. As in previous studies, the relationship is strong if yield measurement is based on self-reported production, which is the traditional approach in household surveys to estimate yields. The relationship disappears and even reverses, however, if yields are measured with crop cuts, which is a standardized, objective method to measure yields. This suggests that the inverse *plot-size* relationship is an artefact of systematic measurement error in self-reported production.

In theory, this result could also be explained by differences in labor input between large and small plots since crop cuts measure the potential harvest (i.e. just before the harvest) and self-reported production the actual harvest (i.e. after the harvest). If small plots are more completely harvested than large plots, an inverse relation would only be observed for self-reported yields and not for crop cuts. Although there is evidence of an inverse relation between plot size and family labor, the inverse relationship between self-reported production per hectare and plot size remained significant even after controlling for labor input.

This leaves systematic measurement error in self-reported production as the remaining explanation for the finding. No one disputes that self-reported production in household surveys is measured with substantial error for various reasons, including the long recall period. Yet, this error is invariably assumed to be independent of land size and therefore not a causal source of the IR. The inverse *plot-size* productivity relationship can readily be explained by measurement error if the measurement error is positive on small plots and if the errors relative to the true production decrease with plot size. Our findings indeed suggest that production is overreported on small plots and underreported on larger plots relative to production as estimated with crop cuts. Moreover, the systematic bias is substantial. Our best estimates indicate that production needs to be overreported by 40 percent to 100 percent on plots of $100m^2$ and underreported by 10 percent to 35 percent on plots of $3000m^2$ to explain the IR.

Systematic measurement error in self-reported production has important consequences for the analysis and interpretation of household survey data. Our analyses do not, however, reveal why households systematically misreport production. One may speculate that this is related to the recently established fact that farmers overestimate land size on small plots and underestimate it on larger ones (Carletto et al., 2015). This phenomenon explains systematic error in self-reported production if farmers simply report production by multiplying 'estimated' plot size with an estimate of 'average' yields. Respondents indeed often use simple heuristics when reporting on complex issues (Tversky & Kahneman, 1975). A second, potential explanation for systematic measurement error is heaping, which occurs frequently in surveys. Small rounding errors are more of a concern on small plots, where the total production is small, than on larger plots, where the error introduced by rounding is small relative to the total production. This asymmetry between small and large plots causes decreasing measurement error in yields with increasing plot size, which can generate an inverse relation. Once we understand why farmers misreport production, surveys could be redesigned to minimize reporting error and statistical procedures and rules of thumb could be developed to recalibrate self-reported production. This is critical since numerous papers rely on self-reported productivity to study, among other things, the profitability of fertilizers and new technologies or the impact of development programs on raising land productivity.

The rejection of the inverse productivity *plot-size* relationship has two important implications for the inverse *farm-size* productivity relationship.[22] First, the hypothesis of missing markets as the main explanation for the inverse *farm-size* relationship had been

[22] The rejection of the *plot-size* relationship also implies that increasing land fragmentation – i.e. subdividing large plots into smaller units – will not increase production. This policy has never been recommended in practice, although it is a clear implication of the inverse *plot-size* relationship.

discredited because missing markets cannot explain productivity differences across plots within the same household (Assunção & Braido, 2007). Attributing the inverse *plot-size* productivity relationship to measurement error thus reinvigorates the more conventional explanation of missing markets for the inverse *farm-size* productivity relationship. From a policy perspective, this implies that reducing frictions in land, labor and credit markets will increase agricultural output. Second, systematic measurement error in self-reported production might also generate or reinforce the inverse *farm-size* relation. Although this could not be tested with our data, attributing the *farm-size* relationship to measurement error in self-reported productivity would refute a stylized fact and challenge one of the key arguments in favor of small-scale agriculture. After all, small farms might not be more efficient than larger ones.

References

Ali, D. A., & Deininger, K. (2015). Is There a farm size–productivity relationship in African agriculture? Evidence from Rwanda. *Land Economics, 91*(2), 317-343.

Assunção, J. J., & Braido, L. H. B. (2007). Testing Household-Specific Explanations for the Inverse Productivity Relationship. *American Journal of Agricultural Economics, 89*(4), 980-990.

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

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

Barrett, C. B., Bellemare, M. F., & Hou, J. Y. (2010). Reconsidering conventional explanations of the inverse productivity–size relationship. *World Development, 38*(1), 88-97.

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

Benjamin, D., & Brandt, L. (2002). Property rights, labour markets, and efficiency in a transition economy: the case of rural China. *Canadian Journal of Economics/Revue canadienne d'économique, 35*(4), 689-716.

Bhalla, S. S., & Roy, P. (1988). Mis-specification in farm productivity analysis: The role of land quality. *Oxford Economic Papers, 40*(1), 55-73.

Blarel, B., Hazell, P., Place, F., & Quiggin, J. (1992). The economics of farm fragmentation: evidence from Ghana and Rwanda. *The World Bank Economic Review, 6*(2), 233-254.

Carletto, C., Gourlay, S., & Winters, P. (2015). From guesstimates to GPStimates: land area measurement and implications for agricultural analysis. *Journal of African Economies*, ejv011.

Carletto, C., Savastano, S., & Zezza, A. (2013). Fact or artifact: The impact of measurement errors on the farm size–productivity relationship. *Journal of Development Economics, 103*(0), 254-261.

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

Carter, M. R., & Wiebe, K. D. (1990). Access to capital and its impact on agrarian structure and productivity in Kenya. *American Journal of Agricultural Economics, 72*(5), 1146-1150.

Chayanov, A. V. (1926). *AV Chayanov on the theory of peasant economy*: Manchester University Press.

Chen, Z., Huffman, W. E., & Rozelle, S. (2011). Inverse relationship between productivity and farm size: the case of China. *Contemporary Economic Policy, 29*(4), 580-592.

Collier, P., & Dercon, S. (2013). African agriculture in 50 years: smallholders in a rapidly changing world? *World Development*.

Eswaran, M., & Kotwal, A. (1986). Access to capital and agrarian production organisation. *The Economic Journal, 96*(382), 482-498.

Feder, G. (1985). The relation between farm size and farm productivity: The role of family labor, supervision and credit constraints. *Journal of Development Economics, 18*(2), 297-313.

Fermont, A., & Benson, T. (2011). Estimating yield of food crops grown by smallholder farmers. *International Food Policy Research Institute. Washington, DC.*

Ferreira, F. H., Chen, S., Dabalen, A., Dikhanov, Y., Hamadeh, N., Jolliffe, D., . . . Sangraula, P. (2016). A global count of the extreme poor in 2012: data issues, methodology and initial results. *The Journal of Economic Inequality, 14*(2), 141-172.

Heltberg, R. (1998). Rural market imperfections and the farm size— productivity relationship: Evidence from Pakistan. *World Development, 26*(10), 1807-1826.

Irz, X., Lin, L., Thirtle, C., & Wiggins, S. (2001). Agricultural productivity growth and poverty alleviation. *Development policy review, 19*(4), 449-466.

Kimhi, A. (2006). Plot size and maize productivity in Zambia: is there an inverse relationship? *Agricultural Economics, 35*(1), 1-9.

Lamb, R. L. (2003). Inverse productivity: land quality, labor markets, and measurement error. *Journal of Development Economics, 71*(1), 71-95.

Larson, D., Otsuka, K., Matsumoto, T., & Kilic, T. (2014). Should African rural development strategies depend on smallholder farms? An exploration of the inverse-productivity hypothesis. *Agricultural Economics, 45*(3), 355-367.

Pašakarnis, G., & Maliene, V. (2010). Towards sustainable rural development in Central and Eastern Europe: Applying land consolidation. *Land Use Policy, 27*(2), 545-549.

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

Taffesse, A. S., Dorosh, P., & Asrat, S. (2011). Crop production in Ethiopia: regional patterns and trends.

Tversky, A., & Kahneman, D. (1975). Judgment under uncertainty: Heuristics and biases *Utility, probability, and human decision making* (pp. 141-162): Springer.

Verschelde, M., D'Haese, M., Rayp, G., & Vandamme, E. (2013). Challenging small-scale farming: A non-parametric analysis of the (inverse) relationship between farm productivity and farm size in Burundi. *Journal of Agricultural Economics, 64*(2), 319-342.

Appendix: full results of table 3,4,5

Table 3: The inverse plot-size relationship for yields based on self-reporting and crop cuts (full results)

	Self-reported measurement				Crop cuts	
	(1)	(2)	(3)	(4)	(5)	(6)
Log of plot size (m²)	-0.397***	-0.396***	-0.303***	-0.161***	0.104***	0.149***
	(-33.26)	(-36.11)	(-12.65)	(-6.22)	(8.63)	(9.36)
Wave	-0.103***	-0.0978*	-0.101*	-0.0966*	-0.427***	-0.403***
	(-1.90)	(-1.78)	(-1.87)	(-1.69)	(-9.45)	(-8.60)
Plot characteristics						
Log distance to dwelling	0.133***	0.110***	0.150***	0.0925**	-0.0188	-0.0285
	(3.99)	(4.26)	(3.27)	(2.20)	(-0.65)	(-1.01)
Plot Slope (percent)	-0.00238*	0.000461	0.00202	0.00132	-0.000340	-0.000522
	(-1.60)	(0.38)	(0.88)	(0.60)	(-0.24)	(-0.38)
Plot Elevation (m)	-0.000246**	-0.000198*	-0.000262	-0.000148	-0.000328***	-0.000248**
	(-1.78)	(-1.83)	(-1.64)	(-1.04)	(-3.08)	(-2.34)
Plot Potential wetness Index	-0.000669	-0.000929	-0.00625	-0.00684	-0.00882	-0.00952
	(-0.11)	(-0.19)	(-0.61)	(-0.77)	(-1.09)	(-1.25)
Household has land title (no=1)	0.0105	0.0190	0.0193	0.0429	0.0436*	0.0532**
	(0.26)	(0.74)	(0.46)	(1.10)	(1.69)	(2.09)
Crop in pure stand (no=1)	-0.304***	-0.325***	-0.308***	-0.313***	-0.0871	-0.102
	(-3.30)	(-3.34)	(-3.39)	(-3.55)	(-1.09)	(-1.28)
Pure stand (no=1)*Wave 2	0.619***	0.619***	0.457***	0.465***	0.0445	0.0602
	(6.24)	(5.90)	(3.44)	(3.70)	(0.39)	(0.54)
Manure applied (no=1)	0.0456*	0.0191	-0.0129	0.0191	-0.0677*	-0.0573
	(1.31)	(0.68)	(-0.26)	(0.41)	(-1.89)	(-1.60)
Compost applied (no=1)	0.0420	0.0196	-0.116	-0.0952	-0.0455	-0.0223
	(0.94)	(0.48)	(-1.51)	(-1.30)	(-0.68)	(-0.35)
Organic fertilizer (no=1)	-0.0266	0.0663	-0.0841	-0.0799	-0.254**	-0.265**
	(-0.29)	(0.81)	(-0.56)	(-0.58)	(-2.27)	(-2.38)
Field irrigated (no=1)	-0.272***	-0.202**	-0.290***	-0.265**	0.0340	0.0520
	(-3.29)	(-2.15)	(-2.78)	(-2.59)	(0.42)	(0.63)
Log fertilizer (kg/ha)	0.0838***	0.0779***	0.0840***	0.0496***	0.0421***	0.0272***
	(12.09)	(12.12)	(9.94)	(6.12)	(6.21)	(3.98)
Household characteristics						
Asset index		0.682***	0.591***	0.514**	0.319*	0.345*
		(4.22)	(2.93)	(2.57)	(1.72)	(1.66)
Female headed household (no=1)		0.0801***	0.0533	0.0561	0.00490	0.0151
		(3.12)	(1.02)	(1.16)	(0.14)	(0.45)
Age household head		0.000379	-0.00409	-0.00942	-0.00208	-0.00256
		(0.11)	(-0.65)	(-1.57)	(-0.42)	(-0.51)
Age²		0.00000278	0.0000461	0.0000909	0.0000332	0.0000357
		(0.08)	(0.77)	(1.59)	(0.67)	(0.71)
Household head can read and write (no=1)		0.0503**	0.00557	0.00674	-0.0270	-0.0297
		(2.40)	(0.15)	(0.20)	(-0.95)	(-1.11)
Labor input (logs)						
Family labor planting (days/ha)				0.123***		0.0718***
				(6.06)		(4.79)
Hired labor planting (birr/ha)				0.0365***		0.0210*
				(2.64)		(1.89)
Exchange labor planting (days/ha)				0.00427		0.0158**
				(0.43)		(2.16)
Family labor harvest (days/ha)				0.292***		0.0591***
				(12.23)		(3.31)
Hired labor harvest (birr/ha)				0.0918***		0.0158
				(5.74)		(1.39)
Exchange labor (days/ha)				0.105***		0.0353***
				(9.94)		(5.34)
Constant	12.64***	11.93***	12.34***	10.33***	10.13***	9.278***
	(36.59)	(28.08)	(19.90)	(18.36)	(23.98)	(21.31)
Household fixed effects	X					
Enumeration area fixed effects		X	X	X	X	X
Observations	25811	25004	5248	5059	5248	5059
R-squared	0.194	0.201	0.110	0.215	0.120	0.140

Notes: ***, **, * denote statistical significance at the 1%, 5% and 10% levels. T-statistics in parentheses. Errors clustered at the enumeration area.

Table 4: Estimation of the IR for several subsamples (full results)

	Wave1		Wave 2		Maize		Outliers discarded	
	Self-reprorted	Crop cuts	Self-reprorted	Crop cuts	Self-reprorted	Crop cuts	Self-reprorted	Crop cuts
Log plot size (m²)	-0.160***	0.171***	-0.129***	0.138***	-0.439***	0.0595	-0.1000***	0.110***
	(-3.67)	(7.78)	(-4.33)	(7.34)	(-7.31)	(1.50)	(-4.91)	(8.27)
Wave					-0.0609	-0.456***	-0.0707	-0.337***
					(-0.51)	(-6.62)	(-1.54)	(-8.73)
Household characteristics								
Asset index	0.250	0.0880	0.771**	-0.105	-0.499	-0.144	0.104	0.0933
	(0.75)	(0.39)	(2.03)	(-0.43)	(-0.72)	(-0.38)	(0.64)	(0.55)
Female headed household (no=1)	-0.0172	-0.00957	0.0750	0.0148	0.0584	0.0436	0.0315	0.0175
	(-0.23)	(-0.17)	(1.38)	(0.43)	(0.53)	(0.56)	(0.86)	(0.73)
Age household head	-0.00419	-0.00150	-0.0130	0.00258	0.00300	0.00159	-0.00985*	-0.00450
	(-0.49)	(-0.22)	(-1.61)	(0.50)	(0.20)	(0.17)	(-1.92)	(-1.22)
Age²	0.0000583	0.0000176	0.000114	-0.00000937	0.0000218	-0.00000383	0.0000932*	0.0000466
	(0.73)	(0.27)	(1.44)	(-0.18)	(0.15)	(-0.04)	(1.87)	(1.26)
Literate household head (no=1)	0.0331	-0.0477	-0.0272	-0.0413	-0.0569	0.0340	0.00537	-0.00649
	(0.62)	(-1.17)	(-0.63)	(-1.33)	(-0.62)	(0.50)	(0.18)	(-0.33)
Plot characteristics								
Log distance field to dwelling	0.0390	0.00130	0.108**	-0.0162	0.0471	-0.136**	0.0623*	-0.0295
	(0.57)	(0.03)	(2.38)	(-0.45)	(0.44)	(-1.98)	(1.73)	(-1.28)
Plot Slope (percent)	-0.00335	-0.00131	0.00481**	0.00111	-0.000374	-0.00299	0.00220	-0.00000410
	(-0.93)	(-0.56)	(2.09)	(0.74)	(-0.06)	(-0.78)	(1.39)	(-0.00)
Plot Elevation (m)	-0.000207	-0.000140	-0.000184	-0.000309**	0.000635**	0.0000369	-0.0000240	-0.000255***
	(-1.00)	(-0.96)	(-1.07)	(-2.32)	(2.22)	(0.16)	(-0.21)	(-2.87)
Plot Potential wetness Index	-0.00922	-0.0224**	-0.00235	-0.00438	-0.0137	0.00202	-0.00148	-0.00655
	(-0.70)	(-2.24)	(-0.20)	(-0.45)	(-0.55)	(0.10)	(-0.21)	(-1.19)
Household has land title (no=1)	0.0198	0.00973	0.0509	0.0834***	-0.000884	-0.0473	0.00568	0.000353
	(0.29)	(0.23)	(0.99)	(2.62)	(-0.01)	(-0.75)	(0.17)	(0.02)
Crop in pure stand (no=1)	-0.322***	-0.137*	0.156**	-0.0459	-0.415*	-0.0750	-0.169**	-0.0793
	(-3.37)	(-1.96)	(2.28)	(-0.76)	(-1.87)	(-0.38)	(-2.43)	(-1.27)
Pure stand (no=1)*Wave 2					0.590**	-0.159	0.282***	0.0316
					(2.41)	(-0.64)	(2.83)	(0.37)
Manure applied (no=1)	0.0933	-0.0719	-0.0456	-0.0127	-0.0413	-0.0382	0.0172	-0.0258
	(1.19)	(-1.16)	(-0.83)	(-0.33)	(-0.39)	(-0.58)	(0.46)	(-0.82)
Compost applied (no=1)	-0.0658	-0.0818	-0.0639	0.0269	0.247	0.219**	-0.125**	-0.0434
	(-0.49)	(-0.85)	(-0.73)	(0.47)	(1.46)	(2.13)	(-2.30)	(-0.72)
Organic fertilizer (no=1)	-0.0433	-0.238**	-0.0726	-0.248	-0.398**	-0.159	-0.114	-0.182*
	(-0.34)	(-2.26)	(-0.19)	(-1.07)	(-2.01)	(-0.97)	(-1.16)	(-1.81)
Field irrigated (no=1)	-0.0136	0.228**	-0.323**	-0.104	-0.291	-0.00751	-0.208**	0.0958
	(-0.09)	(2.02)	(-2.13)	(-1.11)	(-0.86)	(-0.03)	(-2.43)	(1.20)
Log fertilizer (kg/ha)	0.0651***	0.0178	0.0326***	0.0318***	0.0335	0.00282	0.0375***	0.0190***
	(4.63)	(1.43)	(3.81)	(5.10)	(1.43)	(0.20)	(5.55)	(3.68)
Soil quality			-0.0635**	-0.0570***				
			(-2.10)	(-2.92)				
Labor input (logs)								
Family labor planting (days/ha)	0.104***	0.0545**	0.181***	0.0934***	0.0679	0.0259	0.0965***	0.0659***
	(3.07)	(2.50)	(8.08)	(5.45)	(1.48)	(0.97)	(5.68)	(5.74)
Hired labor planting (birr/ha)	0.0404*	0.0243	0.0464***	0.0243**	-0.000233	0.0432***	0.0240**	0.0176*
	(1.66)	(1.21)	(2.89)	(2.17)	(-0.01)	(2.76)	(2.20)	(1.93)
Exchange labor planting (days/ha)	0.0293*	0.0200	-0.0120	0.000367	-0.0256	-0.0108	0.00940	0.0117**
	(1.94)	(1.51)	(-0.88)	(0.04)	(-0.99)	(-0.56)	(1.10)	(2.12)
Family labor harvest (days/ha)	0.268***	0.0769***	0.347***	0.0696***	0.241***	0.0226	0.218***	0.0389***
	(6.20)	(3.15)	(12.19)	(3.78)	(4.65)	(0.68)	(11.43)	(2.63)
Hired labor harvest (birr/ha)	0.127***	0.0389**	0.0804***	0.0114	0.0547	0.0132	0.0693***	0.0130
	(4.36)	(2.19)	(3.63)	(0.75)	(0.82)	(0.45)	(5.03)	(1.21)
Exchange labor (days/ha)	0.0914***	0.0295***	0.110***	0.0435***	0.0827***	0.0344**	0.0769***	0.0278***
	(5.71)	(3.01)	(8.39)	(5.39)	(2.69)	(2.06)	(9.18)	(4.94)
Constant	9.749***	8.631***	9.594***	8.593***	9.812***	6.990***	9.874***	9.363***
	(11.67)	(15.15)	(9.98)	(13.73)	(9.36)	(8.41)	(24.04)	(24.50)
Observations	1860	1860	3152	3152	726	722	4664	4557
R-squared	0.198	0.101	0.249	0.094	0.356	0.182	0.168	0.133

Notes: ***, **, * denote statistical significance at the 1%, 5% and 10% levels. T-statistics in parentheses. Errors clustered at the enumeration area. Enumeration area fixed effects are included.

Table 5: Estimation of the inverse relation between labor input per hectare and plot size (full results)

	Family labor planting	Hired labor planting	Family labor harvest	Hired labor harvest
Log plot size (m²)	-0.382***	0.0661***	-0.311***	0.0769***
	(-40.18)	(8.30)	(-22.08)	(8.01)
Wave	0.122*	0.0476	-0.153**	0.0875**
	(1.94)	(1.54)	(-2.12)	(2.17)
Plot characteristics				
Log of distance field to dwelling	0.0550**	0.0358	-0.00727	0.0498
	(2.22)	(1.31)	(-0.21)	(1.59)
Plot Slope (percent)	-0.00185	0.00113	0.000395	-0.000670
	(-1.58)	(0.90)	(0.27)	(-0.57)
Plot Elevation (m)	-0.0000239	0.000112	-0.0000934	0.0000797
	(-0.29)	(1.42)	(-0.80)	(0.78)
Plot Potential wetness Index	-0.000877	0.000700	0.00470	-0.0000857
	(-0.20)	(0.14)	(0.74)	(-0.01)
Household has land title (no=1)	-0.00814	-0.0131	0.0160	-0.0737**
	(-0.36)	(-0.49)	(0.49)	(-2.30)
Crop in pure stand (no=1)	0.00280	0.0314	0.311***	0.0426
	(0.03)	(0.54)	(2.69)	(0.87)
Pure stand (no=1)*Wave 2	-0.165*	-0.0740	-0.240**	-0.0706
	(-1.70)	(-1.14)	(-2.04)	(-1.24)
Manure applied (no=1)	-0.0582**	0.0319	-0.191***	-0.00558
	(-2.24)	(1.58)	(-5.24)	(-0.18)
Compost applied (no=1)	0.0689*	-0.0556	-0.155***	0.00302
	(1.92)	(-0.92)	(-2.73)	(0.09)
Organic fertilizer (no=1)	-0.0956	-0.0142	-0.0430	0.0790
	(-1.22)	(-0.24)	(-0.50)	(0.78)
Field irrigated (no=1)	-0.00977	-0.00528	-0.0444	-0.161*
	(-0.15)	(-0.14)	(-0.45)	(-1.96)
Log fertilizer (kg/ha)	0.0264***	0.0410***	0.0993***	0.0463***
	(4.83)	(5.64)	(15.12)	(5.22)
Household characteristics				
Asset index	0.126	1.279***	-0.177	0.992***
	(0.54)	(5.46)	(-0.67)	(4.47)
Female headed household (no=1)	0.118***	-0.103***	0.185***	-0.151***
	(4.43)	(-3.37)	(3.83)	(-3.43)
Age household head	0.0249***	-0.0123***	0.0220***	-0.0181***
	(6.02)	(-2.76)	(4.08)	(-3.68)
Age²	-0.000238***	0.000121***	-0.000219***	0.000175***
	(-5.85)	(2.78)	(-3.84)	(3.53)
Household head can read and write (no=1)	0.0598***	-0.0565**	0.0402	-0.0465*
	(2.64)	(-2.24)	(1.30)	(-1.88)
Constant	3.365***	-0.127	3.873***	0.271
	(9.31)	(-0.45)	(8.79)	(0.74)
Observations	24743	24743	25004	25004
R-squared	0.299	0.038	0.169	0.035

Notes: ***, **, * denote statistical significance at the 1%, 5% and 10% levels. T-statistics in parentheses. Errors clustered at the enumeration area.