

Using Global Position System for Land Measurement: Testing the Farm size-Productivity Relationship¹

Sara Savastano, University of Rome Tor Vergata
Calogero Carletto, The World Bank
Alberto Zezza, The World Bank

WORK IN PROGRESS

1. Introduction

There is a well-established consensus that agriculture statistics are plagued with different types of non-sampling errors which may derive from distortions introduced via data collection methodology and techniques, interviewer's effect, respondents' interpretation of specific questions and their motivation to provide accurate answers, as well as respondents' characteristics like their level of education and gender. A relatively recent addition to the toolbox of farm and household survey practitioners is the use of Global Position System (GPS) devices for the measurement of land areas. Household surveys, particularly multi-purpose surveys like the Living Standards Measurement Study (LSMS), commonly rely on farmers' self-reporting to estimate the area under ownership and cultivation. As the technology become more affordable and its reliability improves, GPS devices are increasingly been used for land area data collection in household surveys. Preliminary empirical evidence (Goldstein and Udry, 1999) suggests that the differences with self-reporting may be substantial, and that such difference varies by farm size. If this intuition is correct, using GPS may have considerable implication on the much debated and contentious relationship between farm size and productivity.

The long-standing controversy over the existence of an inverse relationship between farm size and productivity (IR henceforth) in developing countries has been recently reviewed by Eastwood et al. (2008) and, particularly in relation to African agriculture, questioned by Collier and Dercon (2009) who maintain that "there are (only) a handful of reasonably careful studies showing the inverse farm-size/productivity relationship in African settings ... but also some showing the reverse (i.e. positive) farm-size/productivity relationship". A substantial part of the debate, and increasingly so in recent contributions, has focussed on whether the IR may be a statistical artefact, stemming from problems with the available data. The debate has also emphasized the possible role of the omission or imprecise measurement of land quality traits in determining the empirical findings on the IR (Benjamin, 1995; Bhalla and Roy, 1998; Lamb, 2003; Barrett et al., 2010). Disentangling the factors driving the

¹ The views expressed in this paper are the authors' only and should not be attributed to the institutions they are affiliated with.

observed relationship is crucial for putting in place the right policies and guide future investments in the agricultural sector in general, and smallholder agriculture in particular. The empirical evidence, however, remains scant. We are aware of only one study (Lamb, 2003) that attempts to empirically test the robustness of the IR relationship to possible errors in land area measurement using IV methods².

Drawing on data from a nationally representative household survey from Uganda, in which both self-reported and GPS-based plot level information were collected, this paper systematically analyzes the difference in land area data using both measurements, and discusses the impact of such differences on estimates of agricultural productivity. In particular, we explore three different issues, namely (i) what are the determinants of the errors in land measurement, (ii) analyze the size the error according to plot size, and landholding, (iii) testing the existence of an inverse farm size relationship, and the extent to which land measurement error affect the relative advantage of small holders implied by an IR.

The paper is organized as follows. The next section succinctly reviews the main strands of literature this paper relates to. Section 3 provides a description of the data. The econometric models used are sketched in Section 4 and the results discussed in Section 5. The final section discusses the contributions of the analysis to the literature and the policy discussion.

2. The inverse farm size-productivity relationship: agronomic dominance or statistical artefact?

The relationship between farm size and productivity in developing countries is one of the longest standing and most contentious empirical issues in agriculture and development economics. The reason for such attention is evident, as much of the current development paradigm and resulting investments have often assumed the existence of an inverse relationship. Several potential factors have been put forth in the literature as possible explanation for such relationship. However, from a policy perspective, the real challenge has been not only to show the existence of an IR but, more importantly, to understand the reasons behind it. Whether the relationship is the result of a host of factors related to differences in market and production processes across farm types or simply land measurement errors has radically different policy implications and, as such, has attracted the attention of many researchers.

² Lamb (2003) attributes differences in fixed and random effects estimates to land measurement issues, but does not have explicit data to account for land measurement errors. Using a profit equation the IR persists in the fixed effect model. In the random effects (less subject to measurement error) the important result is that the coefficient of land is exactly equal to 1 so that the IR is completely explained by the combination of land quality and market imperfections.

Starting with the seminal work of Sen (1962, 1966) who observed an inverse relationship between farm size and output per hectare in Indian agriculture, a large number of empirical studies have presented evidence that appears to corroborate that hypothesis (Barrett, 1996; Carter, 1984, Benjamin and Brandt, 2002; Berry and Cline, 1979). A smaller set of empirical studies however does not find evidence of such a relationship (Hill, 1972, Kevane, 1996, Zaibet and Dunn, 1998). Eastwood et al. (2010) provide a careful discussion of both the theory and the empirics of the IR debate, a full review of which is beyond the scope of this paper.

It will suffice here to note that, following Barrett et al. (2009), an inverse relation between farm productivity and farm size may have three main explanations: (i) imperfect factor markets, (ii) omitted variables and, in particular, omitted controls for land quality, and (iii) statistical issues related to the measurement of plot size. Imperfect factor markets (labour, land, insurance) are linked to differences in the shadow price of production factors that in turn lead to differences in the application of inputs per unit of land, in ways that are correlated with farm size. Much of the earlier contributions to the IR debate focussed on testing this type of explanation. Later studies (e.g. Bhalla and Roy, 1988; Benjamin, 1995) have challenged the existence of the IR based on the observation that when land quality controls are introduced in the analysis, the strength of the IR often diminishes substantially or vanishes altogether. Barrett et al. (2010) utilize a dataset that includes laboratory measures of soil testing to conclude that in fact only a very limited proportion of the IR can be explained by differences in land quality. Lastly, attention has been drawn to the possibility that the existence of the IR may be a statistical artefact deriving by measurement error in land data (Lamb, 2003). A similar explanation is also given by Barrett et al (2010), after failing to explain the observed IR otherwise.

However, for the IR to be partially or fully explained by errors in land measurements, smaller farmers would have to systematically underreport land area with respect to larger farmers, thus resulting in artificially inflated yields in the bottom part of the distribution. However, as reported by De Groote and Traorè (2005), small farmers tend to overestimate their land holding – while large farmers tend to underreport – hence increasing the likelihood of analysts finding a (spurious) and even stronger inverse relationship in empirical studies based on more accurate measures.

Land area measurement is one of the fundamental components of agricultural statistics, and it is therefore not surprising that the interest of agricultural statisticians in the possibility of applying technological innovations such as satellite imagery and GPS devices to land area measurement is growing exponentially with the increasing affordability and precision of these technologies and their applications. Kelly et al. (1995) have long identified the use of GPS has having the potential to contribute to making land area measurement a much less costly and time consuming exercise than traditional methods.

Keita and Carfagna (2009) provide a discussion of the performance of different GPS devices compared to traditional methods (rope and compass), which they consider the ‘gold standard’. They conclude that the GPS is a reliable alternative to traditional measures (80 percent of the plots in their sample is measured with negligible error), but that on average GPS measures tend to underestimate plot area somewhat. They find one main reason of error in GPS measures to be the density of plot tree canopy cover and to some degree to weather conditions at the time of measurement. The advantage of their analysis is to test the use of the GPS under several hypothesis, namely, the influence of plot size on the accuracy of the measurements with GPS receivers, time needed for measurement with traditional methods versus GPS, stability of measurement of GPS, repetition of measurements and accuracy, accuracy by different type of GPS receivers. The evidence presented seems to prove that GPS devices allow to measure farm size with enough accuracy compared to traditional and objective land measurement methods such as compass and meters.

In a study comparing crop area measurement via actual physical measurement with farmers’ own assessment following enumerator assisted visual inspections, De Groote and Traorè (2005) on the contrary find evidence that the bias in land self-reporting by farmers varies with plot size. In particular they find that small farmers, those owning less than one hectare of land, tend to over-estimate farm size, whereas larger farmers tend to under-estimate land ownership, with the extent of the bias negatively correlated to the size of the landholdings. Also, they find the size of the bias to vary across plots depending on the crop planted. In a somewhat different application in the context of Peruvian market access self-reported data *vis a vis* “true travel time” using GPS, Escobal and Laszlo (2008) show how if the error is correlated with the main outcomes of interest the conclusion of the analysis may be biased and driven by a spurious correlation in that the deviation between the ‘true’ travel time and the respondents’ estimate is determined by observable socio-economic variables related to the outcome of choice. Gibson and Mckenzie (2007) show evidence of the non-random distribution of measurement error in several of the distance and location studies they review.

Data source and descriptive statistics

Our analysis is based on the Uganda National Household Survey (UNHS) round implemented by the Uganda Bureau of Statistics (UBOS) in 2005-2006. The UNHS is a standard multi-purpose household survey with a sample of some 7,500 households out of which some 5,500 are located in rural areas and cultivating land. One unique feature of this survey that makes our analysis possible is that it contains plot level information on agricultural land area measured through both GPS and farmers’ own estimates.

The total number of plots with self estimated measure is 13,959. About 65 percent of these plots (9173 plots) have a corresponding GPS measure. However, to control for total households landholding, we have restricted our sample to 5767 plots for which we have complete information on

both type of measures on all plots cultivated by farmers. This will allow us to match the information with the agriculture production section which contains information at the household/farm level. We end up with a sample of 2893 rural households for which we have both non-zero land area measures for the entire households' landholdings.³

Table 1 reports some summary statistics for the plots for which we have information on both type of area measurements, along with the mean and standard deviation (in acres) of plot size measured through GPS, self-reporting, and the absolute and relative difference between the two measures. For simplicity of presentation, we will take the GPS measure as the benchmark and talk of farmers over- (or under-) reporting plot size whenever the area self-reported is larger (smaller) than the area measured by GPS. We do acknowledge however that the GPS measure may be subjected to a certain degree of inaccuracy (Keita and Carfagna, 2009).

On average the two methods produce strikingly similar estimates of land area. The average size of plots using GPS is 2.24 acres, a mere 0.11 acres larger than the area reported by farmers. The sample level means however mask pervasive differences in measurement that emerge at closer scrutiny. In the overall sample, farmers overestimate plot size in 54.12 percent of the cases (or 3,121 plots), and underestimate it in 44.13 percent of plots (equivalent to 2545 plots). For the remaining 101 plots (1.75 percent) the survey reports identical measures with either method. For the plots where a 'positive discrepancy' is observed, namely a GPS measure larger than self reporting, the amount of area overestimated is larger in absolute value (1.07 acres on average) than for the plots for which the bias is negative (-0.67 acres).

Table 2 summarizes the size of the bias by deciles of landholding area computed using the GPS measure. The last column of table 2 reports the discrepancy in percentage terms, in order to account for scale effect. In percentage terms, the magnitude of the discrepancy appears to increase monotonically as one moves from the bottom (smallest landholders, --142 percent) to the top deciles (largest landholders, 21 percent). The majority of the average negative discrepancy is due to difference in measures in the first 2 deciles. The sixth deciles, for example, account 2 times less than the average (-16 percent against --27 percent). The size of the positive discrepancy increases three times between the eight and nine deciles (between 0.13 and 0.39 acres), and is the largest in the top deciles (3.96 acres).

All this suggests that there is a pattern in the direction of the bias, with smaller farmers generally over-reporting their land relatively more than larger farmers, and with the largest farmers actually under-reporting land size somewhat (but less in relative terms). Up to 4.57 acres (the 7th deciles), the bias is on average negative, in the top three deciles, the opposite is true.

³ Using the overall sample of plots, the average number of plots cultivated by farmers is 3.36. When we restrict the sample to the 5,767 plots with GPS measure, we obtain an average number of 2.64 plots per household.

As in Keita and Carfagna (2009) our data point to the fact that the bias is much more pronounced in the tails, thereby confirming that the majority of the bias lies there.

What is particularly interesting about the distribution of these data, is that they seem to go contrary to the results of De Groote and Traorè (2005) and the expectations of Barret et al. (2009). As we recalled earlier, De Groote and Traorè find that farmers have a tendency to underestimate small plots (which they define as below 1 hectare⁴) compared to larger ones, whereas the opposite is true in our case. Barrett et al. (2010: p. 89) in fact only provide an example of what the impact of measurement error would be if it was a statistical artefact stemming from land measurement error, but it is relevant to our discussion later in the paper, that their example assumes that error is negatively correlated with farm size, i.e. that “respondents with smaller plots and farms systematically *over-report* the size of their farm. What we find, is that small farms in the Uganda sample (defined as farmers with less than 4.57 acres – equivalent to 1.8 ha) are actually more likely to under-report the size of their farms. Also, the GPS and self-reported plot area measures in our sample display a correlation of 0.96 (, well above the 0.15 found by Goldstein and Udry in their Ghana dataset.

One additional issue in our data, that is immediately apparent from the visual inspection of the distributions of the two land measures in Figure 2, is the considerable tendency of respondents (or enumerators) to round their reported plot size to nearest acre or half acre. This ‘heaping’ in the response patterns is not uncommon (Roberts and Brewer, 2001) but we suspect it may be particularly important in the case of land measurement since it is bound to matter proportionally more to the left of the distribution, as the same amount of rounding represents a larger percentage of the actual plot size.

Figure 1 also confirms the observation made above based on Table 1. While the two distributions are overall quite close (and in fact the means are not much different) at any single point in the distribution they deviate considerably, in a way that appears to be considerably influenced by heaping in the self-reporting distribution as opposed to a smooth curve for the GPS measure. Finally, the comparison of the two distributions appears to support the case for treating the GPS measure as the more accurate of the two.

This, as any other systematic pattern in land measurement error, has the potential to introduce a bias in the estimation of agricultural/land productivity. To explore that point descriptively, Figure 3 draws the relationship between deciles of cultivated land and farm yields (computed as the ratio of the value of agricultural production over farm size). The slope of both lines is negative, pointing to an inverse yield-farm size relationship in both cases, but the line is much steeper when the GPS land measure is used.

⁴ 1 hectare is equal to 2.47 acres.

Table 3 summarizes yield computed through GPS and area self-reported. We divide farmers between small and large, and we take the 5th deciles as the threshold. Small farms, those cultivating landholding smaller than 2.3 acres exhibit higher yields when area cultivated is measured through GPS as compared to self-reporting. In both cases, smaller farmers have higher yields than larger ones.

The inclusion of a more accurate measure of land area in this dataset seems to be strengthening the empirical case for the existence of an inverse farm size productivity relationship, rather than weakening it. This is consistent with what we expected after looking at the distribution of the discrepancies in the two measures, but appears to contradict expectations based for instance on the analysis in Lamb (2003), who finds that the IR disappears after introducing random fixed effects which he hypothesize capture the influence of measurement error. Nor it would be related to difference in by factor market imperfections or unobserved soil quality as reviewed recently by Barrett et al. (2010)

3. The econometric approach

In order to deepen the analysis of (a) the characteristics and determinants of the discrepancy between our two land measures, and (b) the implications of using one or the other on the IR hypothesis, we estimate the two models specified below.

In order to identify what are the factors affecting the plot bias between the two measures, we propose to estimate the following function, at the plot level:

$$e_i = \beta x_i + u_i \quad (1)$$

Where e_i is the plot size specific bias, x_i is a (K+1)-row vector of control variables with '1' as its first element, $\beta = (\beta_0, \beta_1, \dots, \beta_K)'$ is vector of parameters to be estimated. u_i is a two-sided error term representing white noise that is assumed to be normally distributed with mean 0 and variance σ_v^2 .

The set of controls included in this regression includes a set of characteristics of the household head (age, education, gender) to proxy respondents' characteristics that are deemed to influence the ability to accurately report land size. We also include plot size, and its squared term, to test whether the negative relationship observed in the descriptive analysis holds in a multivariate framework, and a dummy reflecting whether the self-reported land variable is a round number, to capture the impact of rounding. Finally we include information on whether the household is involved in disputes over land ownership, as we expect such households to have less information or interest in these plots.

Next we estimate a standard model for testing the existence of the IR, and to understand how land measurement errors weaken, eliminate or reinforce this empirical evidence. Our model is based on the one originally proposed by Binswanger, Deininger and Feder (1995), and not dissimilar from the approach used by several others including, most recently, Barrett et al. (2010). We estimate the following function:

$$\ln \frac{Y_i}{A_i} = \beta_0 + \beta_1 \ln A_i + \beta_2 X_i + \beta_3 B + \beta_4 R + u_i$$

(2)

where Y_i is the logarithm of the net revenue per acre⁵, A_i the total area operated, X_i denotes a vector of households' characteristics that influence production such as the availability of family labor, the gender, age and education of the household's head, value of inputs used, and a set of land quality variables, B is the bias in both measurements, R is the rounding effect, and u_i is the error term. We estimate two versions of this relationship, one using GPS and the other one self-reported land measures, and compare the two to gauge the impact of measurement on the IR, which is captured by the coefficient on the land size variable. We also run a quintile regression of the same model, to investigate whether the coefficient on the land variable may change at different points in the distribution.

A coefficient below one on the land area variable indicates the presence of the IR. The further below 1 the coefficient, the stronger is the IR.

4. Econometric results

The results of the model (1) regression are reported in Table 4. As it emerged already by the descriptive analysis in section 3, our data show a positive association between plot size and the error in measurement in levels. The discrepancy (the error) increase with farm size with larger farms having the tendency of under-reporting their land, and small farms tend to overestimate. As already stated, this results is contrary to De Groote and Taoré (2005) hypothesis. The inflection point (see the negative coefficient on the square term) kicks in at very large plot sizes (around 170-200 acres, depending on the specification).

The other signs are also as expected. The presence of rounding contributes to increasing measurement error, as we hypothesized based on the heaping displayed by the self-reported distribution in Figure 2. While the education and gender of the household head variables are not significant, his/her age is: older respondents are likely to be less accurate in their reporting of plot

⁵ We also tested with the gross revenue per acre, but did not find any significant difference, we therefore presents results for the net revenue per acre.

sizes. Household who are involved in disputes over land are also likely to make larger errors in self-reporting, which we hypothesize is due to the fact that such disputes diminish their interest, access, and in general knowledge about plot characteristics. Contrary to our expectation plot demarcation does not seem to reduce the bias in self-estimating land. The coefficient of the dummy if the plot has a fence is positive and significant thereby suggesting an increase in the bias⁶.

Table 5 presents the results of the net revenue regression in model (2). In columns (1) and (2) land is measured with the traditional respondent self-report, whereas columns (3) and (4) GPS measure of land size is used. The main variable of interest in this table is the (log) of land size, as it is this coefficient that captures the size-productivity relationship. A negative coefficient indicates an inverse relationships. In all the specifications, with both communities and enumerators fixed effects, the estimates support the IR hypothesis. When the GPS measure is used, however, the magnitude of the coefficient increases significantly approaching -1, indicating an even stronger IR;

This result goes against the hypothesis according to which the IR would be a statistical artefact due to small farmers supposedly under-reporting their farm size. In our sample, small farmers in fact over-report land size more than larger farmer do, and it is the very large famers who are actually more likely to under-estimate their holdings, thus resulting in artificially higher yields. As a result, when the more accurate GPS measure is used instead of the famer's self-reported one, the estimated slope of the function becomes steeper indicating an even stronger IR⁷.

To control for land characteristics we have used variables related to soil quality, topography, and irrigation. Whereas land with good soil quality and hilly land affect positively farms 'profit, irrigation does not appear to be a main factor helping at increasing farm outcome. The variable is significant but negative only when controlling for enumerators fixed effects. We have also introduced household specific controlled variables. As expected older household head have larger profits, but higher levels of education of the head does not seem to affect agriculture net profits. Female households heads experiences lower levels of agriculture gains.

⁶ The regression has been repeated introducing plot quality dummies such as soil quality, irrigation and topography. However, the results are not significant and do not change the overall specification of the model. .

⁷ To look further into this relationship we have re-estimated it in a quintile regression framework to assess whether the slope changes across the land size. We find no evidence of that. The coefficient on the plot size variable is significant in all quintiles and its magnitude does not change by any appreciable amount.

5. Conclusions

According to the analysis in this paper, two conclusions seemingly supported with some degree of general validity by recent contributions to the literature, are to be questioned or revisited. At a minimum, they do not apply to the study of Ugandan agriculture.

Firstly, we conclude that in Uganda small farmers tend to over-report plot size relative to relatively larger farmers, while the largest farm groups under-report plot size on average. This is contrary to the common perception according to which small farmers would under-report land size, whereas large farmers would do the opposite.

Secondly, we find that the empirical validity of the IR hypothesis is strengthened, not weakened, by the availability of better measures of land size collected using GPS devices. Lamb (2003) concludes that by controlling for measurement error in land size, the statistical case for the IR disappears. Barrett et al. (2010) on the other hand, show that including rigorous controls for land *quality* the IR does not significantly affect the findings concerning the IR. Along the same lines this paper argues that introducing rigorous controls for land *quantity* does not affect the evidence concerning the IR, which is in fact strengthened when the better land data are used⁸.

Self-reported measures of land size are notoriously imprecise. In large household surveys in developing countries they have however for a long time been the only option available to practically collect data on the physical dimension of the plots owned or cultivated by the household. More recently, however, the greater availability of more affordable and more reliable GPS devices has made GPS measurement a practical alternative that is increasingly being applied in surveys worldwide.

Being able to measure land with any degree of accuracy is clearly of outmost importance in economies that are largely agricultural based and for communities that derive a large share of their livelihood from agriculture and for whom land constitutes the main, when not the only, capital asset. In particular, an accurate measure of land size is necessary if one is to measure agricultural productivity with any degree of confidence.

⁸ We also use land quality controls, but these are based on self report while Barrett's are based on laboratory soil testing.

This paper has shown that GPS measures of land quality are a valid alternative to self-reports in large household surveys. While our explanation of what causes error in self-reporting is not fully satisfactory, we do find however that self-reported land measures are a reasonable alternative in a well conducted survey. The overall distribution, mean and standard deviation of the two distributions are fairly close to each other. The respondents' rounding of responses can however create fairly serious measurement error particularly when plots are small, as is the case here (but also in most of the developing regions).

6. References

- Barrett, C. B. (1996). On price risk and the inverse farm size–productivity relationship. *Journal of Development Economics*, 51(2), 193–215.
- Barrett, C., Bellemare, M.F., & Hou, J.Y. (2010). Reconsidering Conventional Explanations of the Inverse Productivity–Size Relationship. *World Development* Vol. 38, No. 1, pp. 88–97, 2010
- Benjamin, D. (1995) Can Unobserved Land Quality Explain The Inverse Productivity Relationship? *Journal of Development Economics* 46 (February) :51-84 .
- Benjamin, D., & Brandt, L. (2002). Property rights, labor markets, and efficiency in a transition economy: The case of rural China. *Canadian Journal of Economics*, 35(4), 689–716.
- Berry, R. A., & Cline, W. R. (1979). *Agrarian structure and productivity in developing countries*. Baltimore: Johns Hopkins University Press.
- Bhalla, S. S., & Roy, P. (1988). Mis-specification in farm productivity analysis: The role of land quality. *Oxford Economic Papers*, 40, 55–73.
- Carter, M. R. (1984). Identification of the inverse relationship between farm size and productivity: An empirical analysis of peasant agricultural production. *Oxford Economic Papers*, 36, 131–145.
- Collier, P. & Dercon S. (2009). African Agriculture In 50 Years: Smallholders In A Rapidly Changing World? Paper presented at the Expert Meeting on How to Feed the World in 2050 Food and Agriculture Organization of the United Nations Economic and Social Development Department, Rome, June 24-26.
- De Groote, H., & Traoré O. (2005). The cost of accuracy in crop area estimation. *Agricultural Systems* 84:21-38
- Escoball, J. & Laszlo, S. (2008). Measurement Error in Access to Markets. *Oxford Bulletin of Economics and Statistics* Volume 70(2), 209–243, April
- Gibson, J. & McKenzie, D. (2007). Using the global positioning system in household surveys for better economics and better policy. Policy Research Working Paper Series 4195, The World Bank.
- Goldstein, M., & Udry C. (1999). Agricultural innovation and risk management in Ghana. Unpublished, Final report to IFPRI.
- Hill, P. (1972). *Rural Hausa: A village and a setting*. Cambridge: Cambridge University Press.
- Keita N., Carfagna E. (2009) Use of modern geo-positioning devices in agricultural censuses and surveys, *Bulletin of the International Statistical Institute*, the 57th Session, 2009, Proceedings, Special Topics Contributed Paper Meetings (STCPM22) organized by Naman Keita (FAO) “Using advanced data collection methods and modern tools to improve agricultural statistics data quality”, Durban, August 16-22, 2009
- Kelly, V., B. Diagana, T. Reardon, M. Gaye, and E. Crawford. (1995). *Cash Crop and Foodgrain Productivity in Senegal: Historical View, New Survey Evidence, and Policy Implications*. MSU Staff Paper No. 95-05. East Lansing: Michigan State University
- Kevane, M. (1996). Agrarian structure and agricultural practice. Typology and application to Western Sudan. *American Journal of Agricultural Economics*, 78(1), 236–245.

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

Roberts, John M. & Brewer, Devon D. (2001). Measures and tests of heaping in discrete quantitative distributions. *Journal of Applied Statistics*. 28(7):887-896

Sen, A. K. (1962). An aspect of indian agriculture. *Economic Weekly*, 14, 243–266.

Sen, A. K. (1966). Peasants and dualism with or without surplus labor. *Journal of Political Economy*, 74(5), 425–450.

Zaibet, L. T.; Dunn E.G. (1998). Land Tenure, Farm Size, and Rural Market Participation in Developing Countries: The Case of the Tunisian Olive Sector. *Economic Development and Cultural Change*, Vol. 46.

Table 1: Summary of statistics at the plot level

	Nb of Plots	Unit	Mean	Std. Dev.	Min	Max
GPS	5767	Acre	2.24	12.43	0.01	600
Self – Reported	5767	Acre	2.13	12.75	0.01	600
Discrepancy (GPS-Self reported)	5767	Acre	0.11	2.28	-49	45
Negative discrepancy (Negative Error => GPS<Self Reported)	3121	Acre	-0.67	1.82	-49	-0.01
Positive discrepancy (Positive Error => GPS>Self Reported)	2545	Acre	1.07	2.47	0.01	44.83

Source: Authors' calculations based on UNHS 2005

Table 2: Mean farm size size and discrepancy characteristics by deciles of landholding (GPS measure).

Deciles	Area of HH landholding	Nb. of plot per hh	Mean farm area using GPS	Mean farm area using Self-Reported	Farm Discrepancy (GPS-Self Reported)	Discrepancy in % terms
1	0.01-.65	1.70	0.37	0.73	-0.36	-142
2	0.66-1.12	2.33	0.90	1.43	-0.53	-60
3	1.13-1.62	2.40	1.37	1.78	-0.41	-29
4	1.63-2.09	2.70	1.84	2.36	-0.52	-29
5	2.09-2.69	2.94	2.38	2.91	-0.52	-22
6	2.7	2.80	3.04	3.53	-0.48	-16
7	3.44-4.57	2.74	3.96	4.10	-0.14	-4
8	4.59-6.16	2.94	5.31	5.18	0.13	3
9	6.17-9.13	3.20	7.46	7.08	0.39	5
10	9.14-600	3.40	21.03	17.07	3.96	21
Total	0.01-600	2.70	4.75	4.60	0.15	-27

Note: area is expressed in acres, and computed for total household landholding.

Source: Authors' calculations based on UNHS 2005

Table 3: Relation between yield and Farms Size

	Area Landholding	Yield (GPS)	Yield (Self Reported)	Bias in yield (GPS-Self Reported)
Small Farms	0.01 – 2.3	263	201	62
Large Farms	2.4 - 600	114	123	-9

Yield computed in Uganda Shilling per acre

Table 4: Determinants of difference in plot measurement (dependent variable: GPS-self reported plot size, in acres)

	Bias in Level	
	With PSU dummies	With Enumerators dummies
Plot size (GPS)	0.04*** [0.000]	0.07*** [0.000]
Square Plot size (GPS)	-0.0001*** [0.000]	-0.0001*** [0.000]
Rounding in Self Reported	0.26*** [0.001]	0.27*** [0.001]
Dummy: parcel has a fence	0.46*** [0.006]	0.27* [0.082]
Head's age	0.72*** [0.000]	0.33* [0.079]
Head's Education	-0.01 [0.553]	-0.01 [0.476]
Dummy Female Head	0.07 [0.366]	-0.02 [0.778]
Dummy: was/is in dispute with other relatives	0.67*** [0.004]	0.33 [0.153]
Constant	-0.55*** [0.000]	-0.39*** [0.000]
Observations	5767	5723
R-squared	0.23	0.08
Threshold for plot size	173 acres	202 acres

p values in brackets

* significant at 10%; ** significant at 5%; *** significant at 1%

Table 5: Testing the IR hypothesis. Dependent variable Log of Net Revenue per acre

VARIABLES	Area Self Reported		Area GPS	
	(1)	(2)	(3)	(4)
Land characteristics				
Log Land Size	-0.63***	-0.67***	-0.85***	-0.98***
	[0.000]	[0.000]	[0.000]	[0.000]
Dummy Rounding	-0.01	0.03	-0.12*	-0.13**
	[0.896]	[0.603]	[0.053]	[0.016]
Log value of ag. Inputs	0.10***	0.11***	0.08***	0.10***
	[0.000]	[0.000]	[0.000]	[0.000]
Log family labor	0.48***	0.50***	0.42***	0.41***
	[0.000]	[0.000]	[0.000]	[0.000]
Log hired labor	0.02	0.02	0.03	0.03
	[0.250]	[0.258]	[0.145]	[0.124]
Share of land GPS with good soil quality	0.20***	0.26***	0.19***	0.26***
	[0.000]	[0.000]	[0.000]	[0.000]
Share of land GPS hilly	0.19**	0.36***	0.20**	0.33***
	[0.050]	[0.000]	[0.046]	[0.000]
Share of land GPS irrigated	-0.06	-0.15**	-0.1	-0.16***
	[0.326]	[0.013]	[0.117]	[0.008]
Households characteristics				
Log Head's age	0.10*	0.13***	0.04	0.11**
	[0.063]	[0.009]	[0.448]	[0.041]
Log Head's education	-0.03	-0.02	0.01	0.01
	[0.527]	[0.561]	[0.850]	[0.892]
Dummy female head	-0.08*	-0.11***	-0.06	-0.09**
	[0.074]	[0.010]	[0.199]	[0.034]
Constant	1.84***	1.65***	3.15***	3.07***
	[0.000]	[0.000]	[0.000]	[0.000]
Fixed Effects				
<i>Communities</i>	<i>YES</i>	<i>NO</i>	<i>YES</i>	<i>NO</i>
<i>Enumerators</i>	<i>NO</i>	<i>YES</i>	<i>NO</i>	<i>YES</i>
Observations	2,860	2,836	2,860	2,836
R-squared	0.60	0.44	0.63	0.47

p values in brackets

* significant at 10%; ** significant at 5%; *** significant at 1%

Figure 1: Bias in Land Measurement: Rounding Problems

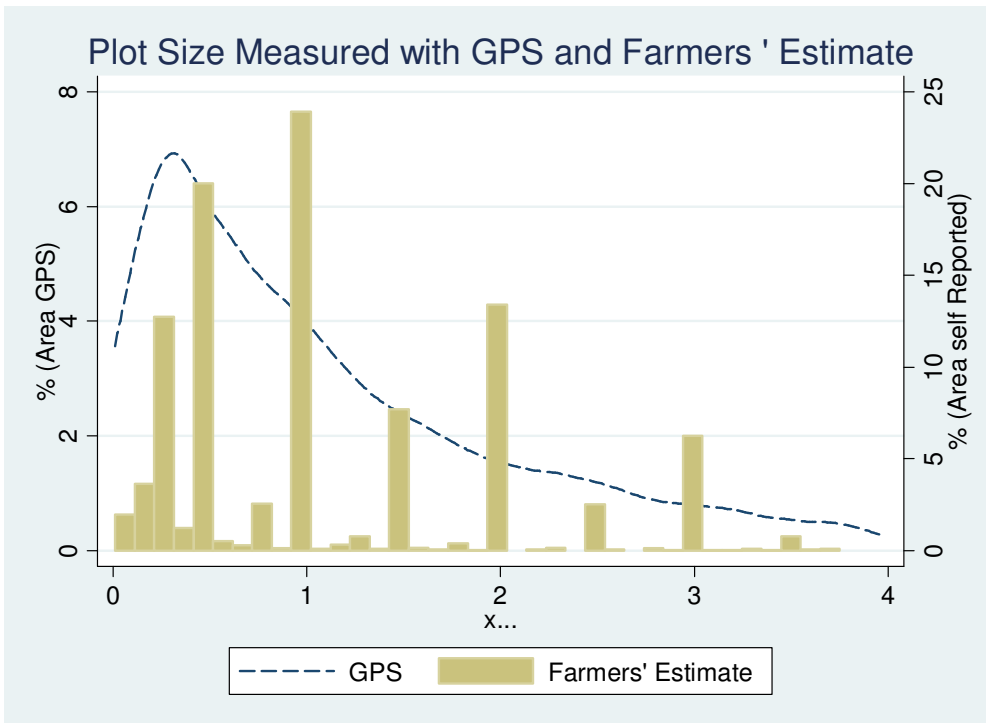
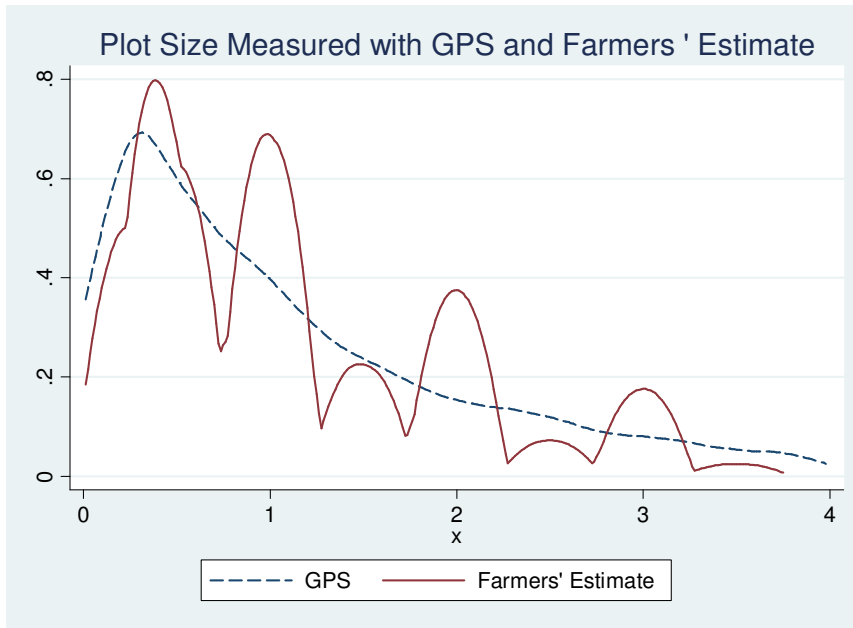


Figure 3: The inverse relationship between yields and farm size: Comparison using GPS and self-reported area estimates

