WPS6550

Policy Research Working Paper 6550

# From Guesstimates to GPStimates

## Land Area Measurement and Implications for Agricultural Analysis

*Calogero Carletto*
*Sydney Gourlay*
*Paul Winters*

The World Bank
Development Research Group
Poverty and Inequality Team
July 2013

Policy Research Working Paper 6550


## Abstract

Land area measurement is a fundamental component of agricultural statistics and analysis. Yet, commonly employed self-reported land area measures used in most analysis are not only potentially measured with error, but these errors may be correlated with agricultural outcomes. Measures employing Global Positioning Systems, on the other hand, while not perfect especially on smaller plots, are likely to provide more precise measures and errors less correlated with agricultural outcomes. This paper uses data from four African countries to compare the use of self-reported and Global Positioning Systems land measures to (1) examine the differences between the measures, (2) identify the sources of the differences, and (3) assess the implications of the different measures on agricultural analysis focusing on the inverse productivity relationship. The results indicate that self-reported land areas systematically differ from Global Positioning Systems land measures and that this difference leads to potentially biased estimates of the relationship between land and productivity.


This paper is a product of the Poverty and Inequality Team, Development Research Group. It is part of a larger effort by the World Bank to provide open access to its research and make a contribution to development policy discussions around the world. Policy Research Working Papers are also posted on the Web at http://econ.worldbank.org. The authors may be contacted at gcarletto@worldbank.org.

# From Guesstimates to GPStimates:
# Land Area Measurement and Implications for Agricultural Analysis[1]

Calogero Carletto, Sydney Gourlay, and Paul Winters[2]

**Keywords**: Survey design; land measurement; GPS measure; inverse productivity relationship
**JEL Classification Codes:** C81, O12, Q12.

---

[1] The views expressed in this paper are the authors' only and should not be attributed to the institutions they are affiliated with. The authors would like to acknowledge comments and inputs from Amparo Palacios-Lopez, Talip Kilic, Gbemisola Oseni, Alemayehu Ambel and Sara Savastano and from participants to the 2013 Annual World Bank Conference on Land and Poverty in Washington, DC.
[2]Carletto and Gourlay are with the Development Research Group of the World Bank. Winters is Associate Professor in the Department of Economics American University, Washington, D.C.

**From Guesstimates to GPStimates:**

**Land Area Measurement and Implications for Agricultural Analysis**

Land area measurement is a fundamental component of agricultural statistics and analysis. Land is a key measure of farmer wealth, a critical input in production, and a key variable for normalizing agricultural input use and output measures. Failure to adequately measure land limits the ability to analyze the agricultural economy. In an econometric analysis, mismeasuring land can potentially lead to questionable estimates of agricultural relationships.

In household surveys, land is commonly measured through farmer self-reported areas. Farmer reported land area is widely used since it can be easily incorporated into a standard household questionnaire and thus requires almost no additional time or money. The use of self-reported areas assumes that farmers are willing and capable of providing reasonably accurate estimates of land area. Evidence from previous studies suggests that land measures taken from farmers are measured with substantial error (Goldstein and Udry, 1999; Carletto, Savastano and Zezza, 2013). For data analysts, the question is one of how much error is associated with a self-reported measure and the implications of this error for conclusions drawn from the analysis of agricultural data.

As Global Positioning Systems (GPS) technology becomes more affordable, accurate and user-friendly, GPS-based area measurement provides a practical alternative to farmer self-reported areas that is increasingly being applied in surveys worldwide.[3] For example, in an assessment of agricultural data collection in Sub-Saharan Africa, Kelly and Donovan (2008) highlight GPS technology as having the potential to enable land area measurement to become a much less time-intensive and costly exercise. Using field experiments, Keita and Carfagna (2009) indicate that 80 percent of the GPS-measured sample plots were measured with negligible error when compared with compass-and-rope estimates, the accepted gold standard. Recent empirical evidence based on the 2005/2006 Uganda National Household Survey (UNHS) comparing GPS-based and self-reported measurement of parcel areas also suggests the existence of systematic errors in self-reported parcel areas (Carletto, Savastano and Zezza, 2013).

Our objectives in this paper are to systematically assess the suitability of GPS devices to replace farmers' self-reporting of plot land area and to determine the implications of using a GPS-based system in agricultural analysis. Comparisons of household self-reported versus GPS-measured distances to market show that not only are self-reported distances characterized by measurement error, but also that the errors are non-classical and dependent on observable socioeconomic variables; this suggests that studies using self-reported measures may be estimating biased effects (Escobal and Laszlo, 2008). Given this is the case, Gibson and McKenzie (2007) argue that GPS can help overcome these problems of measurement in data collection. In the same vein, we demonstrate that using GPS for land area measurement may have extraordinary effects on the analysis of mundane agricultural questions.

[3] The area of an agricultural parcel can also be measured by traversing or delineating parcel boundaries on ortho-corrected remote sensing imagery. Traversing is completed using a tape and compass and is considered the gold standard of land measurement, but its implementation is time-consuming and costly. One Uganda study compared land measurement by traversing versus using GPS units, and found that the average time use per plot measured was over three hours for traversing, which was more than three times as much as the GPS technology (Schoning, Apuuli and Muwanga-Zake, 2005). Traversing is therefore infeasible in the context of national household surveys. Delineating parcel boundaries on satellite imagery is a potentially accurate alternative, but at present largely impractical, particularly in tree-dense areas and areas with regular cloud cover where the ability to make accurate and timely measures is limited. The spatial and temporal extent of national household surveys generally makes the acquisition and processing of such high resolution imagery largely cost prohibitive.

Three steps are taken to meet the objectives of the paper. First, we examine the differences between the two measures. Second, we determine to what degree the source of differences between self-reported area and GPS measures are due to observable factors. Finally, we compare the use of self-reported and GPS land measures in estimating the inverse farm-size productivity relationship to determine the implications for measurement error on agricultural analysis. To that end, we build on Carletto, Savastano and Zezza (2013) to systematically test the existence of an inverse farm size-productivity relationship in a number of different geographic contexts. The choice of the inverse productivity relationship is motivated by the fact it is a controversial relationship widely discussed in the agricultural development literature and land measurement has been noted as a potential issue in estimating the relationship (Lamb, 2003; Barrett Bellemare and Hou, 2010).

Meeting these objectives is possible because of recently collected, multicountry data coming out of the Living Standards Measurement Study—Integrated Surveys on Agriculture (LSMS-ISA). The project has collected relevant information from a number of countries in Sub-Saharan Africa to compare the approaches to land data collection and corresponding agricultural variables in order to assess the implications of such approaches to the analysis of agricultural relationships. We use data from four of the most recently completed surveys—Malawi, Uganda, Tanzania and Niger.

We find that there is a systematic discrepancy between GPS and self-reported plot areas with results indicating that, on average, farmers tend to over-report land area. This is particularly true for farmers at the lower end of the farm size distribution. While rounding of self-reported areas plays an important role in the misreporting as do household and plot level factors, this systematic discrepancy remains, and even strengthens, when controlling for these factors. Exploring the implications of these results on the analysis of the inverse productivity relationship, we find that, while in most countries the relationship slightly weakens, it does not disappear or change sign, thus confirming in each country, as well as in the pooled data, the persistence of an inverse farm size productivity relationship. Further, gradual introduction in the specification of additional control variables creates a greater divergence between the GPS and self-reported estimates indicating a level of endogeneity of the self-reported areas.

The remainder of this paper is organized as follows. Section 2 discusses the issue of land measurement and how it has been addressed in the LSMS-ISA surveys. The section also includes descriptive statistics comparing the measures and how they differ as well as an analysis of the difference between the two measures. Section 3 considers the implications for land measurement on agricultural analyses, particularly the widely analyzed inverse productivity relationship. Section 4 then provides conclusions regarding the implications of the analysis for studies that use self-reported land areas as well as the implications for collecting land data given time and monetary constraints.

## 2. Data and land measurement

Suppose that the GPS measure, $GPS_i$, represents the true area of a plot of land. Self-reported land area, $SR_i$, can then be noted as follows:

$$SR_i = GPS_i + \varepsilon_i \qquad (1)$$

where $\varepsilon_i$ is the error associated with that estimate. Even if classical measurement error can be assumed, where this error ($\varepsilon_i$) in self reporting is considered to be independent of the GPS-level of land area as well as all other variables in an economic model, estimates of economic relationships can be biased (Bound, Brown and Mathiowetz, 2001). In this section, we begin by carefully considering how land area is determined and the implications for the error in this measurement.

There are a number of reasons why self-reported land areas may be subjected to measurement error. First, farmers may knowingly over- or under-state their landholdings for strategic reasons if they perceive information may be used for a certain purpose such as property taxes or access to a program. Second, there are also a number of factors that can lead to unintentional errors in reporting. There is a natural tendency to round off numbers and provide approximations of land areas, which leads to lumping of the data. Geography, particularly parcel slope, can also change the way farmers interpret the land (Keita and Carfagna, 2009). Slope-related effects on area measurement are rooted in the fact that the actual area should be the horizontal projection of the parcel, as opposed to the parcel area itself since plants and trees grow vertically rather than perpendicular to the slope and therefore require for their growth a vertical cylinder of soil (Keita, Carfagna and Mu'Ammar, 2010; Muwanga-Zake, 1985). The difference between actual area and projection appear to be particularly important for slopes greater than 10 degrees (Fremont and Benson, 2011).

Survey design itself can lead to measurement error as well. When farmers are required to report in a single unit, enumerators may be forced to convert the self-reported area from local units if farmers are unfamiliar with the required unit. This can lead to inconsistent conversions across enumerators or data quality issues if the required unit is not used. In some cases, enumerators are advised on common conversion factors, but only for standard units (as opposed to local units) and for common fractions (quarter, half and three-quarters) or they are given conversion estimates in paces for square areas. Alternatively, farmers can report local units and conversion factors can be applied prior to analysis to standardize the units. This use of local units can severely limit reliability of self-reported areas because conversion factors may be unreliable or location specific and acquiring conversion factors for individual zones of the country is difficult to accomplish with accuracy.

This range of factors—intentional or unintentional misreporting, rounding, topography and the existence of local units—opens the door for systematic measurement error in land data based on self-reported areas. It is the presence of these errors that led to a search for alternative measure and particularly the use of GPS. GPS is costly in terms of the units themselves—although this cost continues to decrease and is becoming negligible—and the cost of labor to visit plots and physically walk the perimeter. The question becomes whether these costs are justified by the benefit of more precise land measure.

Despite the great potential of GPS technology, GPS-based coordinates are subject to known types of measurement error rooted in satellite position, signal propagation, and receivers. Approximate contributions of these factors to the overall position error range from 0.5 to 4 meters (Hofmann-Wellenhof, Lichtenegger and Wasle, 2008). On a large plot this may not be substantial, but on a smaller plot, the errors may be significant. The number of satellites, in particular, can cause the distribution of position error to be elliptical, rather than spherical (van Diggelen, 2007). Although position estimates are subject to a certain level of inaccuracy and may be distributed in a non-spherical manner, in theory the error associated with area measurement should be random—that is, the factors that cause non-spherical position error are largely macro level factors that are unlikely to change in the short period of time required to pace the perimeter of a plot, rendering the position error distribution consistent at all points along the perimeter.[4] A study by Bogaert, Delince and Kay (2005) using simulated coordinates and European Geostationary Navigation Overlay Service (EGNOS) augmentation concluded the position error can be reasonably assumed to be

[4] One factor that could possibly change the distribution of the position error around the plot is multipath, or signal reflection. Multipath is typically caused by large buildings, however, and given the rural setting analyzed here, this type of error is assumed negligible.

normally distributed. Overall, from a technical standpoint, GPS-measured areas would be expected to create land data with classical measurement error.

While, technically speaking, GPS area measurement error should be random, the use of the technology in practice has produced different results. For reasons left unexplained by device manuals and technological articles, handheld GPS devices often slightly under report the true area on average with empirical evidence suggesting that the GPS-based area measurement appears to be less precise to a non-trivial degree for plots smaller than 0.5 hectares or 1.24 acres (Schoning, Apuuli and Muwanga-Zake, 2005; Keita and Carfagna, 2009). However, in field experiments using various GPS devices, Hejmanowska, et al (2005) found that Garmin units did not bias estimates either downward or upward in their sample of plots from 0.369 -3.92 hectares although other units did show bias in one direction or the other. The smaller area as measured by GPS could be explained if the GPS unit, and the enumerator using the device, cuts corners, thus shaving off the precise corner point of the plot if there is one, an effect that could be exacerbated by the speed with which the plot is navigated. Analysis by Bogaert, Delince and Kay (2005) confirms the hypothesis that operator speed has an impact on GPS measurement accuracy while adding that the optimal speed is not fixed, but rather varies with plot size. Finally, training of enumerators in the use of the device and the manner in which results are recorded could influence final measurement.

Even with these issues, GPS is considered to be a more accurate method of area measurement than self-reporting. Although the issues with GPS measurement error should not be overlooked and are discussed again in the conclusion of the article, here GPS measurement is considered closer to the true land area and we refer to deviations from the GPS area as self-reported measurement error.

*LSMS-ISA data*

The LSMS-ISA data sets create an opportunity for analyzing the different methods of land data collection since the studies employ GPS measurement while continuing to collect self-reported land areas. Data are currently available for four LSMS-ISA countries: Malawi (2010/11), Uganda (2009/10), Tanzania (2010/11) and Niger (2011). For agricultural households in each of the countries, the plot manager or head of household is first asked to estimate the plot areas and then the plots are measured using GPS, so as not to influence the farmer’s reported figure.

Depending on the survey, farmers may be required to report in a single unit (as in Uganda, Tanzania and Niger) or they may have the choice to report in a number of units (as in Malawi). In either scenario, the estimate is subject to previously noted errors. GPS devices allow for measurement in a variety of units, but are typically recorded in acres in LSMS-ISA projects and rounded to the nearest hundredth.

The exclusion of certain plots from GPS measurement was mandated by the survey manual for the sake of time and budget. In Malawi, for example, enumerators were instructed to measure all plots within two hours of walking from the household. In Tanzania the guidelines were to measure all plots within one hour (via any mode of transportation) and Uganda required all plots within the enumeration area to be measured (Government of Malawi, NSO, 2010; World Bank Group, n.d.; UBOS, 2009). Table 1 provides basic information on the surveys considered in this study and how land data were collected in each case. The number and fraction of total plots or parcels that have both GPS and self-reported areas varies by country. In Niger and Uganda, the survey captured 47% and 58% of plots on GPS while Tanzania and Malawi have more complete data reaching 78% and 96%, respectively. The variation is related both to the rules of data collection and the concentration of agricultural production. For this analysis, the data are limited to households in which all plots were measured with GPS[5] and the validity of the analysis is then solely for these subpopulations.

*Difference between self-reported and GPS land areas*

Looking at the mean GPS and self-reported land area measurements across LSMS-ISA countries gives cause for concern on the equality of the measurement methodologies. Table 2 reports the mean values for the GPS and self-reported plot and farm areas for Malawi, Uganda, Niger and Tanzania as well as a compilation of metadata.[6] In simply looking at the overall means by country, the difference in the two methods is not remarkably distinct—at the plot level it is less than 0.4 acres in all cases with an overall difference of only 1.3% in the combined data. Similar differences for the farm level are found.[7] While differences between self-reported and GPS area (referred to as *bias* in this paper) at both the plot and farm level are found, the divergence in the overall mean values appears small, only rarely reaching over 10% (for the plot and farm level in Tanzania and Uganda).

Disaggregating the data by size—constructed based on levels of GPS measurements—reveals systematic differences in the discrepancy between the two measures, matching the patterns observed by Carletto, Savastano and Zezza (2013) and De Groote and Traore (2005). Smaller plots show gross over reporting of areas by farmers. In the lowest group, the mean self-reported plot areas are over 90% of the average GPS measure for that group in all cases, suggesting systematic over reporting by farmers. The area of large plots, on the other hand, is systematically under reported. In the largest land area group of the metadata, the mean self-reported plot area is more than 30% below the average GPS measure for that group. Tanzania, Niger and Uganda show that the largest plots are under reported by more than 25%, while the largest plots in Malawi are cut short by 59.1%. At the farm level similar patterns emerge with over reporting of land areas relative to GPS measures by smallholders and under reporting by largeholders. In interpreting whether plots and farm sizes are generally under or over reported it should be kept in mind that the number of plots over 5 acres in the metadata are 5% of the total and the number of farms 11% of the total. In most countries, it remains a relatively small fraction of the total with the largest fraction being 31% of plots in Niger. Given this is the case, plot size is generally over reported.

In each of these countries, the bias in measurement begins with extreme highs for small plots and falls as the plot area increases, generally converting to under reporting by farmers in the third or fourth group. This transition from positive to negative bias results in the country average washing out to a passable difference. Even though there are more smallholders, the small positive absolute differences in the plot size average out with the fewer but larger negative absolute differences in plot size reported for large plots. However, differences in means tests confirm that the country averages of GPS and self-reported areas are, in fact, significantly different in all countries except for Niger, where the difference is only significantly different at each disaggregated level. Wilcoxon matched-pairs signed-rank tests (Wilcoxon, 1945) for each country indicate that the distribution of

[5] Probit analysis on GPS reporting shows that distance from the household was a primary factor in the decision to measure. Additional factors (results available in the annex, Table A1) that appear to matter include if the household is urban, the number of household members, land tenure status, and plot size and slope, although with minor magnitude, suggesting that enumerators were less inclined to measure larger, steeper plots.

[6] In order to limit the impact of GPS imprecision at small areas, plots with self-reported or GPS areas less than 0.01 acres were excluded from the analysis. This includes 16 plots from Tanzania, 16 plots from Niger, 47 from Malawi and 2 from Uganda.

[7] Using median values instead of means provides largely similar results. The primary difference is in the lower categories where the percent bias tends to be smaller as the weighting on larger values is reduced. The trends across land categories are the same.

GPS and self-reported areas are significantly different in the pooled data as well as in Malawi and Niger.

Since a key potential component of the error in self-reported land area is rounding, as a first look at the source of mismeasurement Figure 1 provides histograms of the self-reported land data (left hand side) compared to the GPS data (right hand side) including all plots with 5 acres or less. The difference is striking. On both graphs there is clustering on the low end of the scale, which is to be expected given that at least 40% of plots in each country are less than 2 acres. In the self-reported areas, however, there is clear heaping on whole numbers as well as intermediate rounding, especially on the lower end on the quarter, half and three-quarter acre marks. In Malawi, farmers had the option of reporting in acres, hectares or square meters but 99.7% chose to report in acres. In Uganda and Tanzania, acres were the only option, which could lead to rounding on the part of farmers or enumerators. In Niger, square meters were the only available unit and therefore the heaping is not centered on the same acre proportions. While the GPS histograms are markedly smoother, the existence of spikes in Malawi and Tanzania GPS figures suggest there is still some rounding of GPS-measured areas because of survey design and technology.[8]

To assess whether rounding is the source of the trends noted in Table 2, Figure 2 reports the distribution of GPS measures relative to "rounded" self-reported areas for the metadata.[9] For areas such as 0.5 acres, 1 acre, 2 acres and similar higher common rounding values, we check to see the distribution of GPS measures. Near perfect reporting of self-reported measures would lead to (i) a narrow distribution of GPS measures around the self-reported acre value, and (ii) symmetry in the distribution around the self-reported value indicating an equal level of under and over reporting.

Looking at the figures, the distributions for lower estimates appear to include a range of GPS values and the distributions are generally asymmetric. For example, the self-reported 0.5 acre measure has GPS values that primarily range from very small levels up to 1 acre but the distribution appears largely symmetric and in fact 50% under report, 46% over report and 4% are almost exact. Similar results are found for those rounding to 1 acre with a primary range between 0 and 2 acres but largely symmetric. However, for measures rounded at slightly higher amounts (2, 3, 4 and 5 acres) more systematic over reporting is found with nearly two-thirds of households providing estimates over the GPS measure. Generally, the distributions of GPS values display a heavier left tail and wider distribution with increasing self-reported plot size, suggesting that over reporting is more prevalent and self-reported precision declines as the self-reported figure increases. For even higher measures, there are too few observations to create clear distributions but the pattern remains the same with more over reporting of land size than under reporting relative to the GPS measure. In general, rounding appears to induce more over reporting.

*Regression analysis of bias*

To analyze the range of factors that may influence differences in land areas as measured by GPS units and farmer estimates, following Escobal and Laszlo (2008) we turn to regression analysis. Using plot-level data, we estimate the following equation:

$$\varepsilon_i = \alpha + \beta GPS_i + \delta X_i' + \mu_i \qquad (2).$$

[8] In Malawi, the interviewer manual instructs enumerators to record only two digits to the right of the decimal and the questionnaire itself only provides this much space—thus, the ten small spikes seen in Malawi's GPS histogram between 0 and 1 acre. In Uganda, similar instructions led the enumerators to record two spaces to the right of the decimal.

[9] Self-reported areas are considered rounded if they are equal to the following: in meters – 1000, 5000, 10,000, 20,000, 30,000, 40,000, 50,000, 100,000, 200,000 or 500,000; in acres – 0.25, 0.5, 1, 1.5, 2, 2.5, 3, 3.5, 4, 4.5, 5, 10, 15, 20, 25, 30, 40, 50, 100 or 300; or, in hectares – 0.5, 1, 2 or 5.

$\varepsilon_i$ measures the difference between the area measures for plot $i$ $(SR_i - GPS_i)$ and can be viewed as an estimate of *bias* associated with self-reporting land areas. Measuring in this manner makes it possible to assess the factors that determine the difference between measures.

With no other controls, $\alpha$ would simply represent the mean difference between the two measures. If $\alpha$ is not significantly different from zero or very small it would suggest that the measures are similar. Adding a measure of the GPS area controls for bias by GPS reported area. If the coefficient $\beta$ is not significantly different from zero, it suggests that there is no systematic difference between self-reported and GPS-measured land area and the direction of overall bias would depend solely on $\alpha$. If the coefficient is significant and negative, it suggests that as land size increases farmers systematically move towards under reporting land areas while if it is significant and positive it suggests that farmers move towards systematically over reporting land areas. Whether they are under or over reported on average for a given land area would depend on both the value of $\alpha$ and the value of $\beta$. Of course, a nonlinear relationship between GPS area and bias might exist and this is explored. If neither $\alpha$ nor $\beta$ are significantly different from zero it would suggest that self-reported land measures are reasonable substitutes for GPS measures.

Including additional controls provides an assessment of whether other hypothesized factors may influence bias. Further, the inclusion of additional variables may drive out the results on GPS land area ($\beta$) or general mean differences ($\alpha$)—that is, make them insignificantly different from zero—and suggest that while there are differences in farmer estimates and GPS measures, these can be controlled for in any analysis of agricultural relationships. $X_i'$ is then a vector of control variables that are likely to influence self-reporting including those previously hypothesized to influence bias such as factors linked to intentional or unintentional misreporting, rounding, and topography. Since local units were not used in the included countries, this should not be a factor and is not considered. Finally, $\mu_i$ is the error term.

Plot-level data are used in the analysis since this is the point of measure and it is this bias that we are interested in understanding. The equation is estimated for the metadata and then the individual country data. The analysis of the metadata takes a stepwise approach. First, the relationship between bias and GPS-measured land area is considered since this is the key relationship of interest as it forms the basis for understanding if self-reported areas are systematically different from GPS measures. Next, a series of rounding dummies are included first to see if these are important in bias and second to see whether this explains part of the relationship between GPS area and bias. Third, household fixed effects are added to the regression to determine the role household factors play in bias. Since household fixed effects require multiple plots per household, the estimates in this table only include plots in households with multiple plots (15,724 observations). Fourth, plot level factors are added to establish if topography or other plot level factors play a role. Robust standard errors have been used in all estimations to control for heteroskedasticity.

Table 3 reports the results of the four steps with mean values of variables reported in the first column. The results of the first specification show that bias in reporting is declining with plot GPS land area indicating that bias is relatively greater for small land sizes. While the squared term is negative, it is relatively small in the range of interest creating a near linear relationship. With a constant ($\alpha$) equivalent to half an acre and a point estimate on the linear term of -0.3, what this means is that self-reported areas are larger than GPS areas for very small plots but that bias tends to decline beyond this, reflecting the systematic differences found in Table 2. Above half an acre, farmers tend to systematically under report land areas.

Adding in the rounding variables in the second specification, the importance of rounding is clearly shown as all categories of rounding are found to be significant and contribute to bias. Note that with the rounding variables included, the relationship between bias and GPS strengthens substantially in effect creating a steeper estimate. Further, the constant ($\alpha$) increases to 1.25. The results then suggest that bias remains positive up until 2 acres—or equivalently, that controlling for rounding, plots under 2 acres tend to be over reported.

The inclusion of household fixed effects leads to a decline in the land-bias relationship suggesting household factors matter to a degree. Note also that there are changes in the magnitude of the rounding coefficients indicating that rounding is partially linked to household factors. Separate analysis using household variables rather than fixed effects (available from the authors) shows that a number of factors including head education and gender are significant predictors of bias, but, as the fixed effect results show, the systematic bias remains. Looking at the fourth specification, the entry of plot characteristics has limited effect on the overall relationship. Only land tenure is linked to bias and topography does not seem to matter. The results for the inclusion of household and plot factors are important since they suggests that the use of household and plot control variables in regressions assessing agricultural relationships is insufficient to deal with the potential bias in self-reported areas.[10]

The last two columns repeat the analysis of the final specification but limiting the sample to observations where the GPS measure is less than 1 acre and then more than 5 acres. The reason for including these specifications is to check the sensitivity of the results to different parts of the distribution. This is particularly important given the potential issues with using GPS measures for smaller land holdings. The results for the observations under 1 acre show an insignificant linear term but a negative and significant squared term for land area. Combined with the result for the constant term ($\alpha$), we again observe over reporting by farmers for very small plots and a shift towards under reporting with movement towards 1 acre. The results for the observations greater than 5 acres show a larger linear term on the GPS measure when combined with the constant, indicating that on average larger plots are under reported. These tests confirm the widespread over reporting by farmers.

Table 4 repeats the analysis using the fourth specification for each of the individual countries included in the metadata. The country-level results confirm those found in the metadata. Constant terms are between 0.54 and 3.005 (although not always significant) and bias is significant and declining with GPS plot area with linear estimates between -0.405 and -0.755. Estimates on the rounding variables also confirm that rounding plays a significant role in the level of bias.

The analysis of GPS and self-reported plot areas establish systematic and significant discrepancy between the two land measurement approaches. Results indicate that while small plots may be over reported by farmers, larger farmers tend to under report their holdings. These results hold even controlling for other factors including rounding, household factors and plot characteristics. These exist across the countries studied and are likely to hold true in other locations. Overall they suggest that land measurement error is non-classical. The results mirror those of Escobal and Laszlo (2008) who found the same result when comparing self-reported and GPS measured distances to market.

The deviation in land area measurement across methodologies found in these countries could serve to either unravel or reinforce commonly trusted agricultural relationships. The implications of using readily available yet imprecise self-reported land area measurements are explored below through an analysis of the inverse productivity relationship.

[10] To ensure metadata results were not driven solely by Malawi, the country with the largest share of plots, additional metadata analysis was conducted excluding Malawi. Results showed consistent trends (full results available from the authors).

## 3. Estimating the inverse productivity relationship

Production is commonly shown to be negatively correlated with farm area. This inverse productivity relationship has been tried and tested by economists for decades, most recently by Carletto, Savastano and Zezza (2013). While some, such as Bhalla and Roy (1988), have argued that land quality differences fully or near-fully explain the inverse relationship, others, like Lamb (2003) and Barrett, Bellemare and Hou (2010) conclude that land quality does contribute to the relationship, but does not explain it away. However, Lamb (2003) notes that results from fixed and random effects models imply that the land size variable may be subject to measurement error suggesting that part of the inverse relationship may be explained by land mismeasurement. If measurement error is a contributing factor, the use of GPS could ameliorate some of the inverse relationship between productivity and land area. To analyze whether the measurement methodology influences this relationship, descriptive statistics are first examined.

Table 5 provides means for the value of output[11] per farm as well as the value of output per acre using both the GPS and self-reported areas as the basis. The means are disaggregated by land area designations to provide an initial assessment of the inverse productivity relationship. The existence of a basic inverse relationship is clear in the descriptive statistics. While the value of output per farm increases with size, as is expected, the value of output per acre falls. This is generally true for both the GPS-measured land area as well as with the self-reported areas although there are some exceptions within the countries. In Malawi, the mean values for both GPS and self-reported land areas do not explicitly exhibit an inverse productivity relationship with rising productivity for larger plot areas, which may be due to high returns in key crops such as tobacco. Niger also has one category that does not match the pattern although given the sample sizes for smaller areas this may be due to outliers.

While looking at the means, an inverse productivity relationship appears to emerge for both GPS and self-reported areas. With simple mean relationships, however, it is difficult to compare across the two measures and hard to know whether other factors besides land size, which are not controlled for, are driving the relationship. To assess whether the measurement methodology used impacts the strength of the relationship to any significant degree, we turn to regression analysis.

*Analysis of value of output and productivity*

Following the standards for productivity (Carlson, Zilberman, & Miranowski, 1993), to assess the inverse productivity relationship one or two of the following reduced-form empirical relationships are estimated:

$$\ln(Y_i) = \alpha_0 + \alpha_1 \ln(L_i) + \alpha_2 P_i + \alpha_3 H_i + \alpha_4 X_i + \vartheta_r + \epsilon_i \qquad (3)$$

$$\ln\left(\frac{Y_i}{L_i}\right) = \beta_0 + \beta_1 \ln(L_i) + \beta_2 P_i + \beta_3 H_i + \beta_4 \frac{X_i}{L_i} + \vartheta_r + v_i \qquad (4)$$

where $Y_i$ is the value of output on farm "i", $L_i$ is land area in acres, $P_i$ are plot characteristics—including farmer-reported soil type—and control for land quality variation, $H_i$ are household characteristics, $X_i$ are farm inputs, and $\vartheta_r$ are locational (enumeration area) effects that control for local factors not controlled for otherwise (market prices, agroecology, etc.).

[11] Output values are calculated using household-provided conversion factors to kilograms and imputing market prices based on enumeration area medians (with a minimum of three observations per crop and enumeration area to minimize the influence of seasonality and outliers). Unit prices are specific to crop and crop condition where applicable (i.e, fresh, dry, shelled, unshelled, etc). The value of output was separately constructed for some countries using unit-specific prices rather than kilogram conversions to validate the robustness of the output calculations.

In equation (3), $\alpha_1$ measures the elasticity of the value of output with respect to operated land area. If this is estimated to be less than one and significant, it suggests that the value of output rises less quickly than operated land area and therefore the inverse productivity hypothesis holds (Lamb, 2003). If constant returns to scale are assumed, all production factors, including yields and inputs, can be converted to per acre terms and equation (4) estimated. For this equation, the inverse productivity relationship holds if $\beta_1$ is found to be significantly less than zero.

Researchers that estimate this relationship using self-reported land areas assume that land has been adequately measured or that mismeasurement can be controlled for in some manner within the specification. If this is not the case, the estimate may be biased. The direction of bias depends largely on the manner in which self-reported areas deviate from the actual land size. For example, it may be the case that farmers consistently over or under report land areas across the land distribution or, alternatively, that over or under reporting varies with land size. The specific issues in reporting will influence the direction of bias associated with the inverse productivity relationship. The previous analysis of the deviation between self-reported and GPS-measured areas indicates that farmers over report small land areas and under report larger areas suggesting estimates of the inverse productivity relationship will vary between the two measures.

To test this hypothesis, we estimate equations (3) and (4) with GPS and with self-reported land areas to identify the difference. Estimating both equations allows verification of the consistency in the relationship regardless of the approach. In running these equations we are testing two distinct hypotheses. First, whether the inverse productivity relationship holds even when using a GPS land measure. Second, whether the estimate using the GPS land measure differs in a statistical sense from the estimate using self-reported land area.

Along with providing a basic estimate of the productivity-land relationship, as seen in equations (3) and (4) additional variables are included (e.g. $P_i, H_i, X_i, \vartheta_r$) in the specification to check whether any estimated discrepancy in the productivity-land relationship between the two land measures can be controlled for through the inclusion of these additional variables. Towards this end, a stepwise approach is taken to see how the difference between the two measures varies with the addition of controls. Each step includes the addition of key sets of controls noted in the equations. The basic model includes only the land variables with locational (enumeration area) fixed effects to control for differences across location (particularly agroecological differences). The initial analysis focuses on the metadata, which is subsequently confirmed using the individual country data.

Table 6 reports the results for the analysis of value of output (equation 3) using the metadata. Tests of difference (p-values) between the estimates using GPS measured land area and self-reported land area are reported at the bottom of the table. The first estimations, which include no control variables but include enumeration area fixed effects, show that for both land measures the estimated elasticities are substantially lower than one, supporting the inverse productivity relationship. The elasticities for the self-reported areas are lower than for GPS areas—tests show that they are significantly lower—indicating that while the results confirm an inverse productivity relationship, the self-reported area results suggest it is stronger than when using GPS areas. The subsequent regressions incorporate plot level variables, household factors and finally labor and non-labor inputs. The inclusion of these factors, while lowering the elasticities for both measures, reduces the estimate more for self-reported areas. The final results provide support for the inverse productivity relationship and indicate that self-reported area indicates a significantly stronger inverse productivity relationship than estimated by the GPS measure.

Table 7 follows a similar analysis estimating in a stepwise fashion the value of output per acre (equation 4) using the metadata. Again, tests of difference between the estimates using GPS and self-reported areas are reported at the bottom of the table. The results are consistent with Table 6. The estimates for both self-reported and GPS land measures suggest an inverse productivity relationship, but the estimates again diverge with self-reported areas indicating a significantly stronger relationship than GPS.

The inclusion of appropriate control variables in both sets of regressions (Tables 6 and 7) appears to create greater divergence between the two land area measures. First, this suggests that self-reported area is more responsive to the inclusion of these controls indicating this land measure is capturing characteristics of the land operator in addition to the land area. This is consistent with the earlier analysis which showed household- and plot-level factors are significant determinants of bias. It is also consistent with the presence of land measurement error, which is expected to lead to biased estimates (Bound, Brown and Mathiowetz, 2001). Second, it suggests that the inclusion of controls in assessing the inverse productivity relationship may lead to an overestimate of the true relation. At minimum, it indicates that control variables do not help overcome the inherent bias generated by using self-reported land areas.

Of course, as noted earlier in the paper, GPS units potentially underestimate land areas for smaller plots (less than 0.5 hectares or 1.24 acres). If that is the case, the overestimate of the inverse productivity relationship might reflect under reporting of smaller plots by GPS rather than over reporting of self-reported areas. To check this possibility, the two equations are first re-estimated using land area dummies to see if the relationship is greater at lower levels (suggesting this is driving the results). The results (available from the authors) indicate that on the contrary, the differences are greater at larger land areas. Second, the equations are estimated separately for farms under 1 acre and those over 5 acres to see if the estimates vary depending on the observations used. The results (reported in the last columns of Tables 6 and 7) appear consistent and suggest self-reporting of land areas overestimates the inverse productivity relationship at all land sizes.

Finally, Table 8 presents the results for the four individual countries included in the metadata using the final specification. For the individual countries, a few differences emerge from the metadata analysis. First, all estimates suggest an inverse productivity relationship although when using self-reported areas for value of output (equation 3) for Niger, the results are not significant, raising the possibility that self-reported areas include sufficient error to falsely reject the hypothesis of an inverse productivity relationship. In the value of output per acre model, however, the results for Niger are consistent in that the inverse productivity relationship is stronger with farmer self-reported areas. Second, while Malawi and Uganda have trends consistent with what is found in the metadata, the Tanzania results imply that the self-reported land area estimates suggest a weaker inverse productivity relationship compared to using GPS.

## 4. Conclusions

The objectives of this paper are to compare the use of GPS devices to traditional self-reporting of plot land area and to determine the implications of using a GPS-based system for the analysis of agricultural relationships. Particularly, we compare GPS land measurement to self-reported land area in estimating the inverse farm-size productivity relationship. Meeting these objectives is possible because of recently collected LSMS-ISA data which includes GPS and self-reported land areas along with detailed agricultural information from four countries in Sub-Saharan Africa.

The analysis indicates that there is a systematic discrepancy between GPS and self-reported plot areas. The data indicate that small plot areas are generally over reported by farmers while larger plot areas are generally under reported. Given the majority of plots in Sub-Saharan Africa are small, over reporting in this critical area of the land distribution appears to dominate. Of course, in other contexts where the land distribution differs, underreporting may dominate. The source of misreporting seems to be related to rounding as well as a number of household factors. While these factors play an important role in the misreporting, systematic discrepancies between the two measures remain, and even strengthen, when controlling for these factors. This suggests that in assessing agricultural relationships using control variables will not address measurement error. The implications of this study mirror those articulated by Escobal and Laszlo (2008) in their assessment of self-reported access to market or public infrastructure; namely, those who use self-reported land areas in an analysis of agricultural relationships need to address the likelihood that responses are reported with error and that this error is correlated with other household characteristics.

Exploring the implications of these results on the analysis of the inverse productivity relationship, we find that using self-reported land areas to assess the inverse productivity relationship tends to lead to an over-estimate of the relationship. While GPS land area measures still find the inverse productivity relationship exists, in our particular analysis it is not found to be as strong as when self-reported areas are used with estimated coefficients for the two measures being statistically significantly different. Further, the use of control variables in assessing the relationship creates a greater divergence between the GPS and self-reported estimates indicating a level of endogeneity of the self-reported areas. It might be argued that the results are driven by under reporting of land areas using GPS measures, particularly on plots smaller than 0.5 hectares where some studies have shown issues with GPS measures (Schoning et al., 2005; Keita and Carfagna, 2009). However, sensitivity analysis suggests that even for larger land sizes, where the confidence of GPS measure is greater, self-reported land areas still overestimate the inverse productivity relationship.

While the results indicate the inverse productivity relationship is overestimated, the results do not appear to consistently hold across all countries. This suggests that the deviation between self-reported and GPS land areas is context specific depending on circumstances within a country. This makes self-reported areas even more problematic as it becomes unclear what exactly it is measuring and what omitted factors the self-reported measure might be capturing. As such, there will always be the potential for biased estimates of agricultural relationship given the endogeneity of self-reporting. Even with potential errors in GPS measures, it should remain exogenous to household characteristics and a preferred option for analyzing agricultural relationships. The final conclusion then reinforces the arguments made by Gibson and McKenzie (2007) that GPS information should become a routine part of survey data collection since they tend to provide better measure and help to identify causal relationships.

**References**


Barrett, C., Bellemare, M., & Hou, J. (2010). Reconsidering Conventional Explanations of the Inverse Productivity–Size Relationship. World Development 38 (1), 88-97.

Bhalla, S., & Roy, P. (1988). Mis-Specification in Farm Productivity Analysis: The Role of Land Quality. Oxford Economic Papers , New Series 40 (1), 55-73.

Bogaert, P., Delince, J., & Kay, S. (2005). Assessing the Error of Polygonal Area Measurements: A General Formulation With Applications To Agriculture. Measurement, Science and Technology 16, 1170-1178.

Bound, J., Brown, C. & Mathiowetz, N. (2001). Chapter 59: Measurement Error in Survey Data. (J. Heckman and E. Leamer, Eds.) Handbook of Econometrics 5, 3705-3843.

Carletto, G., Savastano, S., & Zezza, A. (2013). Fact or Artefact: The Impact of Measurement Errors on the Farm Size-Productivity Relationship. Forthcoming Journal of Development Economics. http://dx.doi.org/10.1016/j.jdeveco.2013.03.004

Carlson, G., Zilberman, D., & Miranowski, J. (1993). Agricultural and Environmental Resource Economics. New York: Oxford University Press.

DeGroote, H., & Traoré, O. (2005). The cost of accuracy in crop area estimation. Agricultural Systems, 84 (1), 21-38.

Escobal, J. & Laszlo S. (2008). Measurement Error in Access to Markets. Oxford Bulletin of Economics and Statistics 70 (2), 209-243.

Fremont, A., & Benson, T. (2011). Estimating Yield of Food Crops Grown by Smallholder Farmers, A Review in the Uganda Context. IFPRI Discussion Paper 01097.

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, Washington DC: The World Bank.

Government of Malawi, National Statistics Office. (2010). Enumerator Manual for the Agriculture Questionnaire and Fishery Questionnaire.

Goldstein, M. & Udry C. (1999). Agricultural Innovation and Risk Management in Ghana. Unpublished, Finalreport to IFPRI.

Hejmanowska, B., Palm, R., Oszczak, S., & Cieko, A. (2005) Validation of Methods for Measurement of Land Parcel Areas, Final Report. AGH University of Science and Technology. http://mars.jrc.ec.europa.eu/mars/content/download/800/5268/file/ValidationMethodFinalVer2_2.pdf

Hofmann-Wellenhof, B., H. Lichtenegger, E. Wasle. (2008). GNSS – Global Navigation Satellite Systems. Springer-Verlag: New York, NY.

Keita, N., Carfagna, E., and Mu’Ammar, G. (2010). Issues and guidelines for the emerging use of GPS and PDAs in agricultural statistics in developing countries. The Fifth International Conference on Agricultural Statistics; Kampala, Uganda .

Keita, N., E. Carfagna (2009). “Use of modern geo-positioning devices in agricultural censuses and surveys: Use of GPS for crop area measurement.” Bulletin of the International Statistical Institute, the 57th Session, 2009, Proceedings, Special Topics Contributed Paper Meetings (STCPM22), Durban.

Kelly, V. and C. Donovan (2008). “Agricultural Statistics in Sub-Saharan Africa: Differences in Institutional Arrangements and their Impacts on Agricultural Statistics Systems A Synthesis of

Four Country Case Studies." Michigan State University International Development Working Paper No. 95, East Lansing, Michigan.

Lamb, R. L. (2003). Inverse Productivity: Land Quality, Labor Markets, and Measurement Error. Journal of Development Economics 71 (1), 71-95.

Muwanga-Zake, E. S. K. (1985). "Sources of Possible Errors and Biases in Agricultural Statistics in Uganda: A Review." Kampala, Uganda: Institute of Statistics and Applied Economics, Makerere University.

Schoning, P., Apuuli, E., & Muwanga-Zake, E. (2005). Handheld GPS Equipment for Agricultural Statistic Surveys. Experiments on Area Measurements Done During Fieldwork for the Uganda Pilot Census of Agriculture, 2003.

Uganda Bureau of Statistics. (2009). UNPS 2009/10 Interviewer's Manual of Instructions.

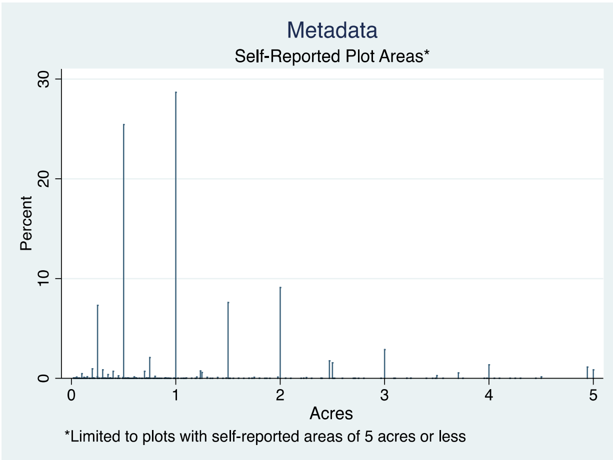
van Diggelen, F. (2007) GNSS Accuracy – Lies, Damn Lies and Statistics. GPS World 18 (1), 1-4.

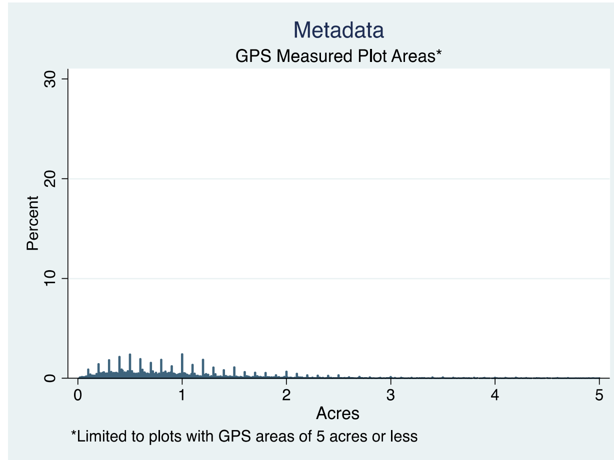
Wilcoxon, F. (1945). Individual comparisons by ranking methods. Biometrics Bulletin 1 (6), 80–83.

World Bank Group. (n.d.). LSMS-ISA Tanzania Country Program. Retrieved 2012, from http://web.worldbank.org/WBSITE/EXTERNAL/EXTDEC/EXTRESEARCH/EXTLSMS/EXTSURAGRI/0,,contentMDK:22938545~pagePK:64168445~piPK:64168309~theSitePK:7420261,00.html

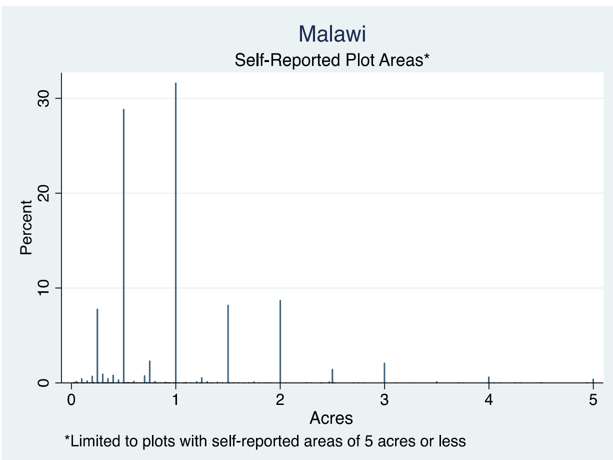
**Figure 1: Distribution of Self-Reported and GPS Plot Area**

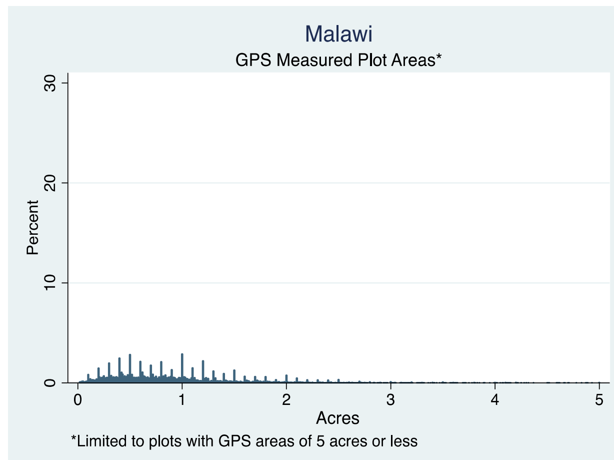
**Metadata**



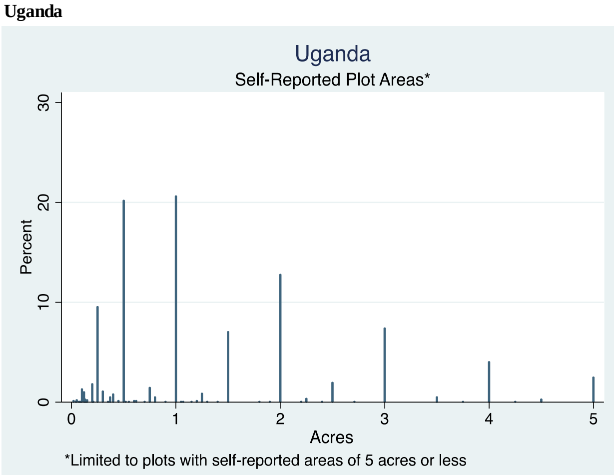



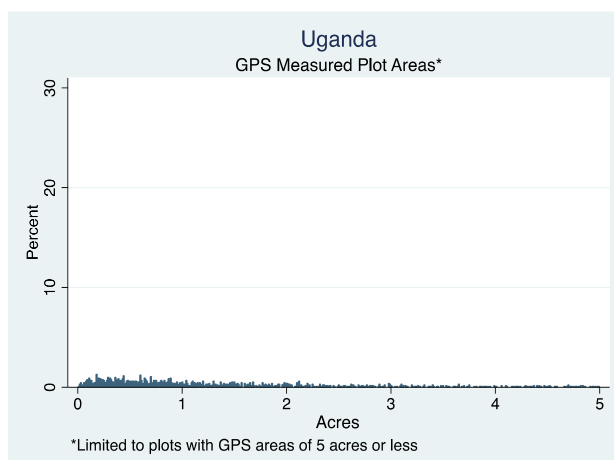

**Malawi**



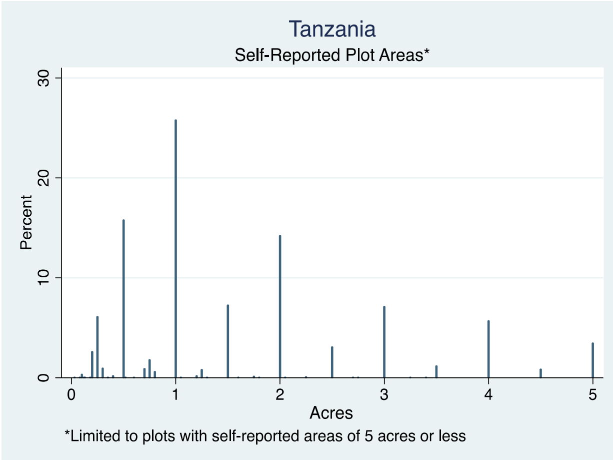



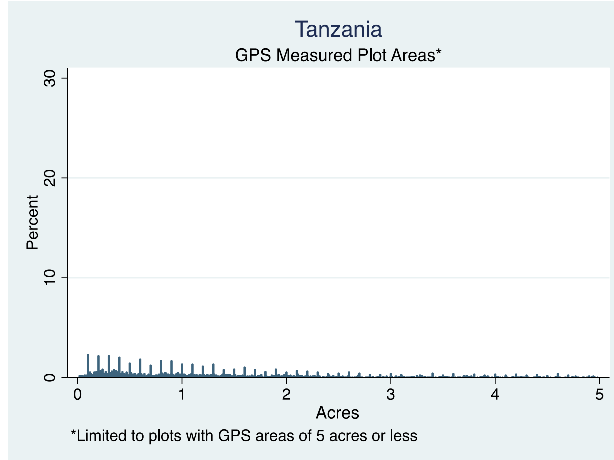

**Uganda**



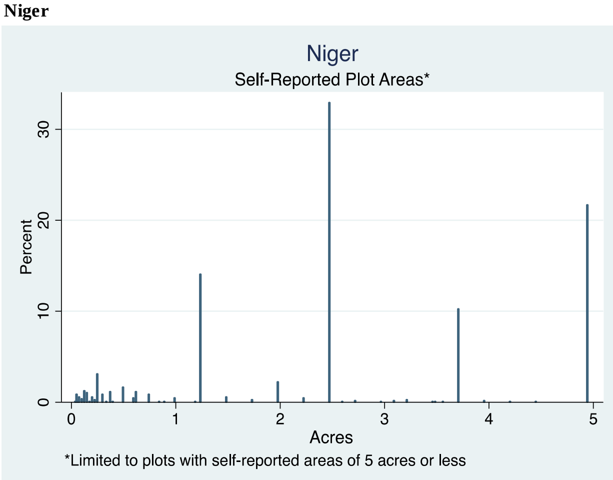



**Tanzania**




**Niger**



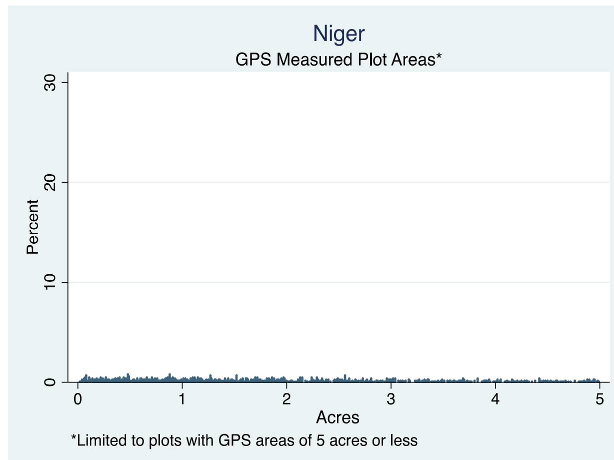
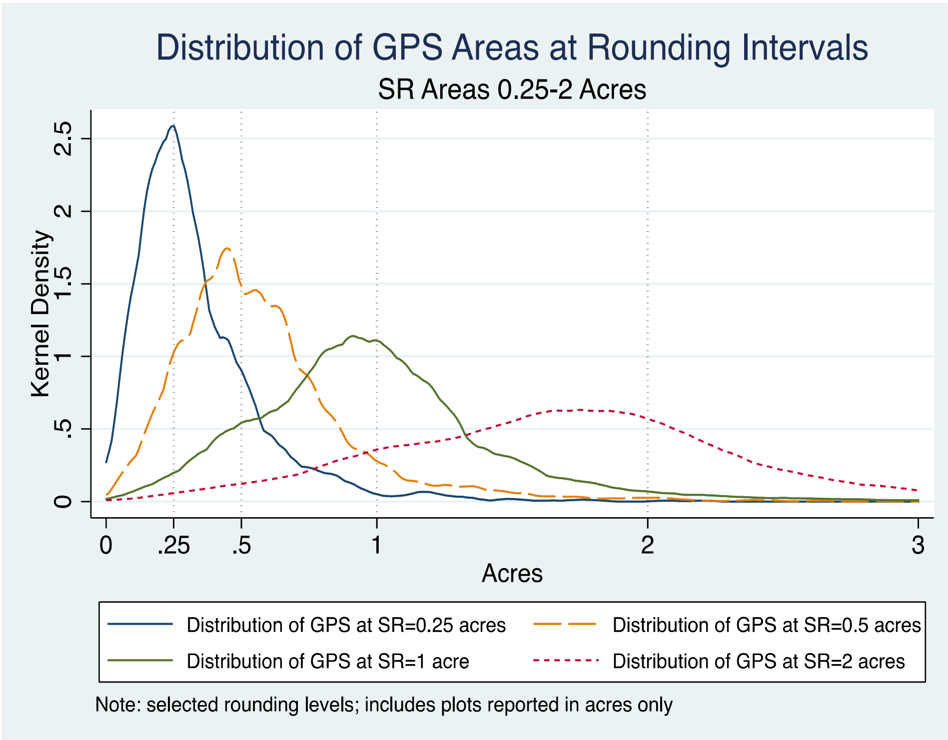



**Figure 2: Distribution of GPS Measurements at Selected SR Areas**

**SR Areas <= 2 Acres:**




**SR Areas 2-5 Acres:**

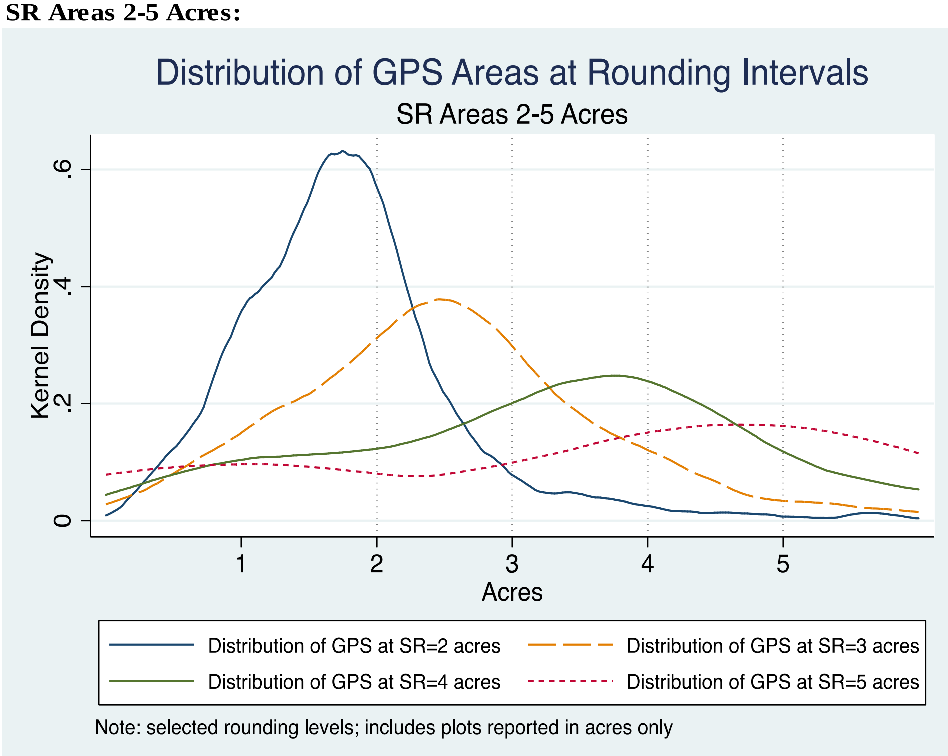


**Table 1: Overview of Data**

| | Malawi | Uganda | Tanzania | Niger |
|---|---|---|---|---|
| GPS unit: | Acres | Acres | Acres | Square Meters |
| Self-reported units: | Square Meters, Hectares, Acres | Acres | Acres | Square Meters |
| Enumerator instructions for converting self-reported areas to required unit when farmer provides an alternate unit: | The enumerator must encourage the use of sq. meters, hectares or acres when possible. If a different unit was used, the following conversions were provided for the enumerator:<br><br>An acre is a measure on the ground of approximately 70 yd x 70 yd or half a standard football field;<br>By casually walking round a square of 50 steps by 50 steps, one covers an area of approximately ¼ or 0.25 acres;<br>An area measuring 22 yd x 22 yd covers 0.1 acres;<br>An area measuring 16 yd x 16 yd covers 0.05 acres;<br>1 hectare =10,000 m2 ≅ 2.5 acres;<br>1 acre ≅ 4000 m2 ≅ 0.4 hectares. | The following conversions were provided for the enumerator:<br><br>An acre is a measure on the ground of approximately 70 yd x 70 yd or half a standard football field;<br>By casually walking round a square of 50 steps by 50 steps, one covers an area of approximately ¼ or 0.25 acres;<br>An area measuring 22 yd x 22 yd covers 0.1 acres;<br>An area measuring 16 yd x 16 yd covers 0.05 acres;<br>1 hectare =10,000 m2 ≅ 2.5 acres;<br>1 acre ≅ 4000 m2 ≅ 0.4 hectares. | Enumerator manual includes conversion factors for square meters, yards and hectares only. No instruction for local units. | None. |
| Guidelines for GPS measurement: | Measurement of plots greater than 2 hours walking distance is decided in consultation with field supervisor. | All parcels located within the enumeration area are to be measured. | Plots within 1 hour of the household by any means of transportation are to be measured. | None. |
| Measurement level: | Plot | Parcel° | Plot | Plot |
| Plots/parcels with GPS and self-reported areas: | 18,281 | 3,372 | 4,717 | 3,102 |
| % of total observations | 96% | 58% | 78% | 47% |

°A plot is defined as a contiguous piece of land within the parcel on which a specific crop or a crop mixture is grown. A parcel may be made up of one or more plots. - UNPS 2009/2010 Interviewer Manual

In Niger, GPS measurements were taken in either or both the post-planting and post-harvest data collection rounds. Farmer self-reported areas were collected only in the post-planting phase, thus the GPS figures reported here reflect only the measurements taken in the post-planting round.

**Table 2: Discrepancy by Plot Size**

| | GPS Area | No. of plots | Plot Mean (acres) GPS Area | SR Area | Mean Bias (SR-GPS) | Bias as % of GPS area | No. of farms | Farm Mean (acres) GPS | SR | Mean Bias (SR-GPS) | Bias as % of GPS area |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **Metadata** | | | | | | | | | | | |
| | < 0.5 | 5332 | 0.30 | 0.60 | 0.30 | 100.0% | 1051 | 0.31 | 0.71 | 0.4 | 129.0% |
| | 0.5 - 0.99 | 6346 | 0.72 | 0.89 | 0.17 | 23.6% | 2250 | 0.74 | 1.11 | 0.37 | 50.0% |
| | 1 - 1.99 | 5856 | 1.35 | 1.43 | 0.07 | 5.2% | 3819 | 1.43 | 1.63 | 0.2 | 14.0% |
| | 2 - 5 | 2446 | 2.90 | 2.88 | -0.02 | -0.7% | 3130 | 3.01 | 3.13 | 0.11 | 3.7% |
| | > 5 | 956 | 11.56 | 7.80 | -3.76 | -32.5% | 1213 | 12.64 | 10.29 | -2.35 | -18.6% |
| **Total** | **0.01 - 97.02** | **20936** | **1.54** | **1.52** | **-0.02** | **-1.3%** | **11463** | **2.81** | **2.77** | **-0.04** | **-1.4%** |
| **Malawi** | | | | | | | | | | | |
| | < 0.5 | 4243 | 0.31 | 0.59 | 0.28 | 90.3% | 837 | 0.32 | 0.68 | 0.36 | 112.5% |
| | 0.5 - 0.99 | 5456 | 0.72 | 0.86 | 0.14 | 19.4% | 1980 | 0.74 | 1.07 | 0.34 | 45.9% |
| | 1 - 1.99 | 4757 | 1.33 | 1.29 | -0.04 | -3.0% | 3371 | 1.42 | 1.56 | 0.13 | 9.2% |
| | 2 - 5 | 1230 | 2.63 | 2.27 | -0.35 | -13.3% | 2354 | 2.94 | 2.94 | 0.00 | 0.0% |
| | > 5 | 69 | 7.98 | 3.26 | -4.72 | -59.1% | 257 | 7.48 | 5.82 | -1.65 | -22.1% |
| **Total** | **0.02 - 25** | **15755** | **0.98** | **1.04** | **0.06** | **6.1%** | **8799** | **1.77** | **1.87** | **0.11** | **6.2%** |
| **Uganda** | | | | | | | | | | | |
| | < 0.5 | 384 | 0.26 | 0.54 | 0.28 | 107.7% | 84 | 0.28 | 0.7 | 0.42 | 150.0% |
| | 0.5 - 0.99 | 321 | 0.73 | 0.93 | 0.19 | 26.0% | 118 | 0.75 | 1.20 | 0.45 | 60.0% |
| | 1 - 1.99 | 314 | 1.43 | 1.49 | 0.05 | 3.5% | 172 | 1.54 | 1.72 | 0.17 | 11.0% |
| | 2 - 5 | 297 | 3.09 | 2.88 | -0.21 | -6.8% | 256 | 3.23 | 3.35 | 0.12 | 3.7% |
| | > 5 | 151 | 13.34 | 9.79 | -3.55 | -26.6% | 201 | 12.38 | 9.62 | -2.76 | -22.3% |
| **Total** | **0.01 - 63.75** | **1467** | **2.41** | **2.17** | **-0.25** | **-10.4%** | **831** | **4.25** | **3.81** | **-0.44** | **-10.4%** |
| **Tanzania** | | | | | | | | | | | |
| | < 0.5 | 554 | 0.26 | 0.63 | 0.37 | 142.3% | 110 | 0.26 | 0.75 | 0.48 | 184.6% |
| | 0.5 - 0.99 | 420 | 0.74 | 1.02 | 0.28 | 37.8% | 138 | 0.77 | 1.20 | 0.43 | 55.8% |
| | 1 - 1.99 | 529 | 1.43 | 1.60 | 0.17 | 11.9% | 239 | 1.49 | 2.07 | 0.58 | 38.9% |
| | 2 - 5 | 508 | 3.16 | 3.01 | -0.15 | -4.7% | 371 | 3.33 | 3.43 | 0.10 | 3.0% |
| | > 5 | 296 | 11.02 | 7.44 | -3.58 | -32.5% | 344 | 12.46 | 9.55 | -2.91 | -23.4% |
| **Total** | **0.02 - 97.02** | **2307** | **2.67** | **2.35** | **-0.32** | **-12.0%** | **1202** | **5.08** | **4.47** | **-0.61** | **-12.0%** |
| **Niger** | | | | | | | | | | | |
| | < 0.5 | 151 | 0.30 | 2.11 | 1.81 | 603.3% | 20 | 0.3 | 2.76 | 2.46 | 820.0% |
| | 0.5 - 0.99 | 149 | 0.78 | 2.70 | 1.93 | 247.4% | 14 | 0.77 | 2.57 | 1.8 | 233.8% |
| | 1 - 1.99 | 256 | 1.49 | 3.46 | 1.97 | 132.2% | 37 | 1.53 | 5.49 | 3.96 | 258.8% |
| | 2 - 5 | 411 | 3.29 | 5.00 | 1.71 | 52.0% | 149 | 3.37 | 6.92 | 3.55 | 105.3% |
| | > 5 | 440 | 11.57 | 8.39 | -3.18 | -27.5% | 411 | 16 | 15.29 | -0.7 | -4.4% |
| **Total** | **0.025 - 84.96** | **1407** | **4.75** | **5.14** | **0.39** | **8.2%** | **631** | **10.86** | **11.74** | **0.89** | **8.2%** |

Note: National data uses sample weights while the metadata does not.
Includes all plots and farms with full GPS measurement.

**Table 3: Determinants of Bias - Metadata**

| | Mean | Bias (SR - GPS) | | | | Bias | Bias |
|---|---|---|---|---|---|---|---|
| *Specification* | | **(1)** | **(2)** | **(3)** | **(4)** | **(4) w/GPS<1** | **(4) w/GPS>5** |
| GPS Plot Area (Acres) | 1.41 | -0.305*** | -0.683*** | -0.583*** | -0.584*** | -0.061 | -0.809*** |
| | | (0.052) | (0.051) | (0.056) | (0.056) | (0.115) | (0.159) |
| GPS Area Squared | 9.30 | -0.005** | -0.001 | -0.004*** | -0.004*** | -0.540*** | -0.001 |
| | | (0.002) | (0.002) | (0.001) | (0.001) | (0.096) | (0.002) |
| *Rounding of SR Areas* | | | | | | | |
| Rounding at areas <1 acre | 0.35 | | -0.982*** | -0.892*** | -0.889*** | -0.166*** | -5.471*** |
| | | | (0.061) | (0.077) | (0.077) | (0.053) | (1.573) |
| Rounding at areas >=1 and <2 acres | 0.34 | | -0.501*** | -0.592*** | -0.583*** | 0.296*** | -7.157*** |
| | | | (0.060) | (0.081) | (0.080) | (0.083) | (1.475) |
| Rounding at areas >=2 and <3 acres | 0.11 | | 0.240*** | -0.105 | -0.097 | 1.218*** | -5.510*** |
| | | | (0.080) | (0.113) | (0.110) | (0.084) | (1.286) |
| Rounding at areas >=3 and <4 acres | 0.03 | | 0.887*** | 0.539*** | 0.547*** | 1.873*** | -3.824*** |
| | | | (0.109) | (0.132) | (0.130) | (0.297) | (1.403) |
| Rounding at areas >=4 and <5 acres | 0.03 | | 1.806*** | 0.695*** | 0.702*** | 2.331*** | -4.274** |
| | | | (0.162) | (0.206) | (0.206) | (0.433) | (1.723) |
| Rounding at areas >=5 and <10 acres | 0.02 | | 3.825*** | 2.091*** | 2.108*** | 4.073*** | -2.768** |
| | | | (0.222) | (0.400) | (0.398) | (0.510) | (1.269) |
| Rounding at areas >=10 acres | 0.01 | | 13.054*** | 10.171*** | 10.181*** | 39.218*** | 2.943 |
| | | | (1.323) | (1.639) | (1.634) | (7.389) | (2.122) |
| *Plot Characteristics* | | | | | | | |
| Plot is irrigated | 0.01 | | | | -0.142 | -0.008 | -10.151*** |
| | | | | | (0.164) | (0.050) | (1.907) |
| Plot has full or partial tree cover | 0.11 | | | | -0.021 | -0.040* | 1.986 |
| | | | | | (0.042) | (0.021) | (1.478) |
| Slope is steep | 0.03 | | | | -0.026 | -0.000 | 0.977 |
| | | | | | (0.096) | (0.035) | (1.302) |
| Soil is sandy | 0.25 | | | | 0.098 | 0.041 | 2.255* |
| | | | | | (0.072) | (0.076) | (1.165) |
| Soil is clay | 0.20 | | | | 0.011 | 0.018 | -0.171 |
| | | | | | (0.045) | (0.027) | (1.296) |
| HH has property rights for plot | 0.10 | | | | 0.733** | 0.007 | 4.367* |
| | | | | | (0.307) | (0.027) | (2.643) |
| Land used for free but not owned | 0.12 | | | | 0.049 | -0.035 | 2.095 |
| | | | | | (0.149) | (0.025) | (1.586) |
| Land rented for a fee | 0.06 | | | | 0.022 | -0.017 | 1.839 |
| | | | | | (0.047) | (0.014) | (1.364) |
| Constant | - | 0.490*** | 1.254*** | 1.265*** | 1.157*** | 0.351*** | 4.559* |
| | | (0.053) | (0.073) | (0.089) | (0.092) | (0.028) | (2.392) |
| Number of observations | 15,724 | 15,724 | 15,724 | 15,724 | 15,724 | 9,226 | 626 |
| $R^2$ | - | 0.323 | 0.565 | 0.609 | 0.610 | 0.715 | 0.788 |
| Household Fixed Effects | - | No | No | Yes | Yes | Yes | Yes |

Notes:
*** p<0.01, ** p<0.05, * p<0.1
Robust standard errors in parenthesis (clustered at fixed effects level where applicable)
Limited to farms with more than one plot
Tree variable is at the farm level for Niger

**Table 4: Determinants of Bias - Country Level**

| | **Malawi** | **Uganda** | **Tanzania** | **Niger** |
|---|---|---|---|---|
| | Bias | Bias | Bias | Bias |
| *Specification* | (4) | (4) | (4) | (4) |
| GPS Plot Area (Acres) | -0.755*** | -0.405*** | -0.461*** | -0.656*** |
| | (0.067) | (0.081) | (0.111) | (0.075) |
| GPS Area Squared | -0.002 | -0.004* | -0.008*** | -0.003*** |
| | (0.014) | (0.002) | (0.003) | (0.001) |
| *Rounding of SR Areas* | | | | |
| Rounding at areas <1 acre | -0.393*** | -0.664*** | -1.324*** | -1.911*** |
| | (0.060) | (0.138) | (0.249) | (0.453) |
| Rounding at areas >=1 and <2 acres | 0.087 | -0.551*** | -1.234*** | -2.715*** |
| | (0.075) | (0.176) | (0.223) | (0.489) |
| Rounding at areas >=2 and <3 acres | 0.848*** | -0.325 | -0.676*** | -2.257*** |
| | (0.127) | (0.272) | (0.210) | (0.430) |
| Rounding at areas >=3 and <4 acres | 1.601*** | -0.058 | -0.312 | - |
| | (0.165) | (0.308) | (0.275) | |
| Rounding at areas >=4 and <5 acres | 2.547*** | 0.647 | 0.222 | -1.012** |
| | (0.185) | (0.427) | (0.336) | (0.479) |
| Rounding at areas >=5 and <10 acres | 3.572*** | 0.343 | 0.457 | 1.272* |
| | (0.213) | (1.000) | (0.510) | (0.703) |
| Rounding at areas >=10 acres | 23.452*** | 3.155*** | 6.967*** | 10.356*** |
| | (7.129) | (0.835) | (1.870) | (2.665) |
| *Plot Characteristics* | | | | |
| Plot is irrigated | 0.024 | 0.308 | -0.761 | 0.414 |
| | (0.044) | (0.714) | (0.482) | (0.799) |
| Plot has full or partial tree cover | -0.023 | -0.106 | -0.000 | - |
| | (0.024) | (0.079) | (0.080) | |
| Slope is steep | -0.004 | -0.036 | -0.050 | 0.179 |
| | (0.035) | (0.253) | (0.261) | (0.840) |
| Soil is sandy | 0.049 | -0.004 | -0.029 | 0.963** |
| | (0.074) | (0.121) | (0.148) | (0.447) |
| Soil is clay | 0.004 | -0.084 | -0.199 | 0.830 |
| | (0.033) | (0.136) | (0.174) | (0.707) |
| HH has property rights for plot | -0.026 | 0.459 | 1.390 | 0.394 |
| | (0.072) | (0.359) | (0.861) | (0.792) |
| Land used for free but not owned | 0.236 | -0.023 | 0.314 | -0.357 |
| | (0.231) | (0.133) | (0.786) | (0.295) |
| Land rented for a fee | 0.017 | -0.090 | 1.330 | -0.194 |
| | (0.049) | (0.122) | (0.831) | (0.396) |
| Constant | 0.691*** | 1.011*** | 0.540 | 3.005*** |
| | (0.045) | (0.152) | (0.828) | (0.426) |
| Number of observations | 11,721 | 1,046 | 1,794 | 1,163 |
| $R^2$ | 0.500 | 0.606 | 0.671 | 0.691 |
| Household Fixed Effects | Yes | Yes | Yes | Yes |

Notes:
*** p<0.01, ** p<0.05, * p<0.1
Robust standard errors in parenthesis (clustered at fixed effects level where applicable)
Limited to farms with more than one plot
Tree variable is at the farm level for Niger

**Table 5: Value of Output by Farm Size**
Weighted Means (USD)

| | GPS Area | | | | Self-Reported Area | | | |
|---|---|---|---|---|---|---|---|---|
| | GPS Farm Area | No. of Farms | Value of Output per Farm | Value of Output per Acre | SR Farm Area | No. of Farms | Value of Output per Farm | Value of Output per Acre |
| **Metadata** | | | | | | | | |
| | < 0.5 | 1051 | 55.4 | 195.68 | < 0.5 | 418 | 62.83 | 314.42 |
| | 0.5 - 0.99 | 2250 | 97.65 | 133.66 | 0.5 - 0.99 | 1667 | 87.45 | 144.87 |
| | 1 - 1.99 | 3819 | 162.13 | 113.66 | 1 - 1.99 | 4107 | 138.53 | 111.47 |
| | 2 - 5 | 3130 | 309.36 | 102.8 | 2 - 5 | 4141 | 279.89 | 98.75 |
| | > 5 | 1213 | 477.96 | 55.74 | > 5 | 1130 | 482.42 | 55.06 |
| **Total** | **0.02 - 200.30** | **11463** | **213.31** | **116.01** | **0.0247 - 252.04** | **11463** | **213.31** | **113.57** |
| **Malawi** | | | | | | | | |
| | < 0.5 | 837 | 52.35 | 185.31 | < 0.5 | 311 | 58.86 | 338.58 |
| | 0.5 - 0.99 | 1980 | 90.41 | 124.4 | 0.5 - 0.99 | 1458 | 79.81 | 134.31 |
| | 1 - 1.99 | 3371 | 152.57 | 107.12 | 1 - 1.99 | 3648 | 131.83 | 106.17 |
| | 2 - 5 | 2354 | 346.39 | 116.19 | 2 - 5 | 3105 | 307.48 | 109.06 |
| | > 5 | 257 | 1045.94 | 146.37 | > 5 | 277 | 893.67 | 123.59 |
| **Total** | **0.02 - 34.18** | **8799** | **210.32** | **122.17** | **0.0247 - 52.1** | **8799** | **210.32** | **121.2** |
| **Uganda** | | | | | | | | |
| | < 0.5 | 84 | 99.23 | 372.92 | < 0.5 | 52 | 86.47 | 390.24 |
| | 0.5 - 0.99 | 118 | 193.37 | 264.44 | 0.5 - 0.99 | 104 | 180.8 | 275.41 |
| | 1 - 1.99 | 172 | 409.42 | 269.75 | 1 - 1.99 | 187 | 263.5 | 218.13 |
| | 2 - 5 | 256 | 421.98 | 125.8 | 2 - 5 | 334 | 433.79 | 147.21 |
| | > 5 | 201 | 1064.33 | 101.13 | > 5 | 154 | 1328.24 | 126.37 |
| **Total** | **0.02 - 73.94** | **831** | **491.78** | **199.01** | **0.08 - 60** | **831** | **491.78** | **194.71** |
| **Tanzania** | | | | | | | | |
| | < 0.5 | 110 | 23.43 | 101.67 | < 0.5 | 37 | 24.84 | 96.41 |
| | 0.5 - 0.99 | 138 | 75.25 | 97.7 | 0.5 - 0.99 | 97 | 55.35 | 88.44 |
| | 1 - 1.99 | 239 | 47.28 | 32.06 | 1 - 1.99 | 245 | 52.11 | 40.53 |
| | 2 - 5 | 371 | 84.04 | 27.28 | 2 - 5 | 538 | 72.81 | 23.62 |
| | > 5 | 344 | 158.5 | 15.95 | > 5 | 285 | 180.01 | 18.06 |
| **Total** | **0.02 - 97.02** | **1202** | **92.24** | **38.36** | **0.1 - 100** | **1202** | **92.24** | **32.58** |
| **Niger** | | | | | | | | |
| | < 0.5 | 20 | 31.76 | 96.31 | < 0.5 | 18 | 71.91 | 219.19 |
| | 0.5 - 0.99 | 14 | 24.31 | 32.92 | 0.5 - 0.99 | 8 | 141.02 | 170.20 |
| | 1 - 1.99 | 37 | 71.32 | 46.78 | 1 - 1.99 | 27 | 58.77 | 43.25 |
| | 2 - 5 | 149 | 108.43 | 34.07 | 2 - 5 | 164 | 95.78 | 29.06 |
| | > 5 | 411 | 221.7 | 16.83 | > 5 | 414 | 210.47 | 16.20 |
| **Total** | **0.0247 - 200.3** | **631** | **171.16** | **26.45** | **0.049 - 252.04** | **631** | **171.16** | **27.23** |

Note: National data uses sample weights while the metadata does not.

**Table 6: Log Value of Output per Farm -- Metadata**

| | Base (with fixed effects) | | Plot Characteristics | | Household Characteristics | | Inputs | | Farms < 1 acre | | Farms > 5 acres | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | GPS | SR | GPS | SR | GPS | SR | GPS | SR | GPS | SR | GPS | SR |
| Log farm size (acres) | 0.802*** | 0.748*** | 0.780*** | 0.723*** | 0.555*** | 0.452*** | 0.370*** | 0.266*** | 0.364*** | 0.127** | 0.316*** | 0.109 |
| | (0.018) | (0.020) | (0.018) | (0.020) | (0.019) | (0.021) | (0.019) | (0.019) | (0.040) | (0.064) | (0.117) | (0.086) |
| *Plot Characteristics* | | | | | | | | | | | | |
| Share of farm with irrigation | | | 0.598*** | 0.517** | 0.536** | 0.492** | 0.393* | 0.357 | 0.389 | 0.564** | 0.998 | 0.340 |
| | | | (0.205) | (0.213) | (0.218) | (0.225) | (0.216) | (0.222) | (0.274) | (0.265) | (1.118) | (0.880) |
| HH has full or partial tree cover | | | 0.508*** | 0.523*** | 0.482*** | 0.490*** | 0.459*** | 0.465*** | 0.515*** | 0.268** | 0.920*** | 0.675*** |
| | | | (0.044) | (0.045) | (0.042) | (0.043) | (0.042) | (0.042) | (0.094) | (0.128) | (0.198) | (0.186) |
| Share of farm with steep slope | | | 0.018 | -0.008 | -0.011 | -0.025 | 0.025 | 0.021 | 0.074 | 0.045 | 0.387 | -0.199 |
| | | | (0.087) | (0.087) | (0.083) | (0.083) | (0.082) | (0.081) | (0.144) | (0.152) | (0.313) | (0.343) |
| Share of farm with sandy soil | | | 0.002 | 0.000 | 0.004 | 0.002 | 0.011 | 0.010 | 0.031 | 0.061 | 0.025 | -0.060 |
| | | | (0.032) | (0.033) | (0.031) | (0.032) | (0.030) | (0.030) | (0.051) | (0.077) | (0.129) | (0.138) |
| Share of farm with clay soil | | | 0.035 | 0.019 | 0.013 | 0.002 | 0.019 | 0.014 | 0.015 | 0.108 | 0.021 | 0.255 |
| | | | (0.032) | (0.032) | (0.030) | (0.031) | (0.028) | (0.029) | (0.056) | (0.073) | (0.223) | (0.212) |
| Share of farm with property rights | | | 0.154* | 0.210*** | 0.192** | 0.236*** | 0.177** | 0.207*** | 0.062 | 0.121 | -0.066 | -0.017 |
| | | | (0.079) | (0.080) | (0.078) | (0.079) | (0.077) | (0.078) | (0.157) | (0.148) | (0.186) | (0.178) |
| Share of farm used free but not owned | | | -0.006 | -0.035 | -0.008 | -0.036 | 0.008 | -0.007 | -0.010 | 0.096 | -0.130 | -0.392** |
| | | | (0.037) | (0.039) | (0.036) | (0.037) | (0.035) | (0.035) | (0.057) | (0.087) | (0.170) | (0.165) |
| Share of farm rented for a fee | | | 0.213*** | 0.177*** | 0.151*** | 0.124** | 0.079 | 0.055 | 0.131* | 0.073 | 0.230 | 0.320 |
| | | | (0.055) | (0.057) | (0.054) | (0.055) | (0.052) | (0.052) | (0.074) | (0.101) | (0.246) | (0.261) |
| *Household Characteristics* | | | | | | | | | | | | |
| Log household head age | | | | | 0.100*** | 0.191*** | 0.023 | 0.075*** | 0.091* | 0.111 | -0.241 | -0.312* |
| | | | | | (0.030) | (0.030) | (0.028) | (0.028) | (0.055) | (0.083) | (0.182) | (0.180) |
| Female HH head | | | | | -0.172*** | -0.188*** | -0.141*** | -0.149*** | -0.110** | -0.145** | -0.137 | -0.376** |
| | | | | | (0.025) | (0.026) | (0.024) | (0.025) | (0.045) | (0.059) | (0.143) | (0.152) |
| Log household head education (years) | | | | | 0.104*** | 0.111*** | 0.053*** | 0.054*** | 0.086*** | 0.086** | -0.115* | -0.135** |
| | | | | | (0.013) | (0.014) | (0.013) | (0.013) | (0.024) | (0.035) | (0.064) | (0.062) |
| Head's primary occupation is agriculture | | | | | 0.059** | 0.087*** | 0.074*** | 0.093*** | 0.032 | 0.165*** | 0.035 | 0.340** |
| | | | | | (0.024) | (0.025) | (0.023) | (0.023) | (0.044) | (0.059) | (0.172) | (0.138) |
| Log number of plots on farm | | | | | 0.605*** | 0.659*** | 0.370*** | 0.387*** | 0.367*** | 0.455*** | 0.318*** | 0.256** |
| | | | | | (0.029) | (0.031) | (0.031) | (0.032) | (0.062) | (0.092) | (0.110) | (0.112) |
| Log HH population | | | | | 0.102*** | 0.152*** | 0.035* | 0.058*** | 0.069* | 0.145*** | 0.126 | 0.159 |
| | | | | | (0.021) | (0.021) | (0.020) | (0.021) | (0.037) | (0.049) | (0.093) | (0.105) |
| *Inputs* | | | | | | | | | | | | |
| Log non-labor inputs value (USD) | | | | | | | 0.222*** | 0.240*** | 0.224*** | 0.237*** | 0.079** | 0.127*** |
| | | | | | | | (0.012) | (0.012) | (0.022) | (0.031) | (0.037) | (0.042) |
| Log family labor days | | | | | | | 0.136*** | 0.164*** | 0.118*** | 0.215*** | 0.086 | 0.216** |
| | | | | | | | (0.016) | (0.016) | (0.029) | (0.038) | (0.092) | (0.084) |
| Log hired labor days | | | | | | | 0.120*** | 0.136*** | 0.114*** | 0.173*** | 0.119*** | 0.141*** |
| | | | | | | | (0.009) | (0.009) | (0.024) | (0.036) | (0.033) | (0.032) |
| Constant | 4.083*** | 4.064*** | 3.967*** | 3.954*** | 3.132*** | 2.691*** | 2.330*** | 1.920*** | 2.241*** | 1.390*** | 3.501*** | 3.414*** |
| | (0.008) | (0.011) | (0.018) | (0.019) | (0.121) | (0.120) | (0.124) | (0.121) | (0.238) | (0.342) | (0.713) | (0.659) |
| Number of observations | 11,463 | 11,463 | 11,463 | 11,463 | 11,463 | 11,463 | 11,463 | 11,463 | 3,301 | 2,085 | 1,213 | 1,130 |
| $R^2$ | 0.258 | 0.214 | 0.273 | 0.230 | 0.330 | 0.299 | 0.389 | 0.373 | 0.259 | 0.265 | 0.204 | 0.242 |
| Fixed Effects | EA | EA | EA | EA | EA | EA | EA | EA | EA | EA | EA | EA |
| Test of difference (GPS and SR; p-values) | 0.0010 | | 0.0004 | | 0.0000 | | 0.0000 | | 0.0013 | | 0.0852 | |

note: *** p<0.01, ** p<0.05, * p<0.1
Robust standard errors in parenthesis, clustered at EA level

**Table 7: Log Value of Output per Acre -- Metadata**
**Farm Level**

| | Base (with fixed effects) | | Plot Characteristics | | Household Characteristics | | Inputs | | Farms < 1 acre | | Farms > 5 acres | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | GPS | SR | GPS | SR | GPS | SR | GPS | SR | GPS | SR | GPS | SR |
| Log farm size (acres) | -0.130*** | -0.194*** | -0.150*** | -0.217*** | -0.367*** | -0.481*** | -0.191*** | -0.251*** | -0.234*** | -0.419*** | -0.273*** | -0.337*** |
| | (0.019) | (0.020) | (0.018) | (0.019) | (0.020) | (0.020) | (0.022) | (0.022) | (0.056) | (0.076) | (0.085) | (0.094) |
| *Plot Characteristics* | | | | | | | | | | | | |
| Share of farm with irrigation | | | 0.584*** | 0.500** | 0.524*** | 0.476** | 0.391** | 0.353* | 0.414 | 0.503** | 0.799 | 0.311 |
| | | | (0.186) | (0.198) | (0.198) | (0.210) | (0.192) | (0.198) | (0.278) | (0.241) | (0.894) | (0.744) |
| HH has full or partial tree cover | | | 0.480*** | 0.486*** | 0.455*** | 0.454*** | 0.426*** | 0.424*** | 0.537*** | 0.313** | 0.631*** | 0.491*** |
| | | | (0.042) | (0.043) | (0.041) | (0.041) | (0.040) | (0.040) | (0.102) | (0.135) | (0.156) | (0.145) |
| Share of farm with steep slope | | | 0.000 | -0.020 | -0.028 | -0.036 | 0.001 | 0.006 | 0.037 | 0.025 | 0.239 | -0.185 |
| | | | (0.086) | (0.083) | (0.082) | (0.079) | (0.081) | (0.077) | (0.169) | (0.169) | (0.225) | (0.216) |
| Share of farm with sandy soil | | | -0.002 | -0.004 | 0.000 | -0.003 | 0.012 | 0.008 | 0.034 | 0.068 | 0.043 | -0.018 |
| | | | (0.031) | (0.032) | (0.030) | (0.031) | (0.028) | (0.029) | (0.054) | (0.082) | (0.104) | (0.115) |
| Share of farm with clay soil | | | 0.027 | 0.011 | 0.006 | -0.006 | 0.013 | 0.008 | -0.000 | 0.107 | 0.063 | 0.222 |
| | | | (0.031) | (0.031) | (0.029) | (0.030) | (0.028) | (0.028) | (0.061) | (0.077) | (0.170) | (0.169) |
| Share of farm with property rights | | | 0.147* | 0.205*** | 0.183** | 0.229*** | 0.153** | 0.183** | 0.066 | 0.114 | -0.088 | -0.058 |
| | | | (0.077) | (0.076) | (0.076) | (0.075) | (0.075) | (0.073) | (0.176) | (0.157) | (0.149) | (0.139) |
| Share of farm used free but not owned | | | -0.013 | -0.038 | -0.014 | -0.039 | 0.005 | -0.007 | -0.027 | 0.101 | -0.062 | -0.319** |
| | | | (0.037) | (0.038) | (0.036) | (0.036) | (0.035) | (0.034) | (0.064) | (0.093) | (0.130) | (0.129) |
| Share of farm rented for a fee | | | 0.211*** | 0.164*** | 0.150*** | 0.111** | 0.075 | 0.044 | 0.136* | 0.090 | 0.105 | 0.209 |
| | | | (0.054) | (0.054) | (0.053) | (0.053) | (0.050) | (0.050) | (0.077) | (0.109) | (0.208) | (0.210) |
| *Household Characteristics* | | | | | | | | | | | | |
| Log household head age | | | | | 0.094*** | 0.191*** | 0.023 | 0.076*** | 0.105* | 0.112 | -0.211 | -0.194 |
| | | | | | (0.029) | (0.029) | (0.028) | (0.027) | (0.060) | (0.091) | (0.130) | (0.142) |
| Female HH head | | | | | -0.169*** | -0.185*** | -0.135*** | -0.142*** | -0.109** | -0.153** | -0.151 | -0.374*** |
| | | | | | (0.026) | (0.026) | (0.024) | (0.025) | (0.051) | (0.064) | (0.108) | (0.119) |
| Log household head education (years) | | | | | 0.102*** | 0.110*** | 0.054*** | 0.054*** | 0.098*** | 0.091** | -0.090* | -0.105** |
| | | | | | (0.013) | (0.013) | (0.012) | (0.013) | (0.026) | (0.037) | (0.053) | (0.053) |
| Head's primary occupation is agriculture | | | | | 0.051** | 0.083*** | 0.070*** | 0.091*** | 0.028 | 0.168*** | -0.029 | 0.233** |
| | | | | | (0.023) | (0.023) | (0.022) | (0.022) | (0.047) | (0.064) | (0.115) | (0.096) |
| Log number of plots on farm | | | | | 0.580*** | 0.634*** | 0.353*** | 0.367*** | 0.379*** | 0.489*** | 0.214** | 0.171* |
| | | | | | (0.028) | (0.029) | (0.029) | (0.030) | (0.065) | (0.097) | (0.090) | (0.092) |
| Log HH population | | | | | 0.101*** | 0.154*** | 0.031 | 0.057*** | 0.087** | 0.163*** | 0.069 | 0.098 |
| | | | | | (0.020) | (0.021) | (0.019) | (0.020) | (0.041) | (0.054) | (0.069) | (0.081) |
| *Inputs* | | | | | | | | | | | | |
| Log non-labor inputs value per acre (USD) | | | | | | | 0.239*** | 0.259*** | 0.206*** | 0.222*** | 0.184*** | 0.219*** |
| | | | | | | | (0.012) | (0.012) | (0.023) | (0.032) | (0.044) | (0.050) |
| Log family labor days per acre | | | | | | | 0.141*** | 0.168*** | 0.122*** | 0.213*** | 0.110 | 0.216*** |
| | | | | | | | (0.016) | (0.016) | (0.031) | (0.039) | (0.075) | (0.074) |
| Log hired labor days per acre | | | | | | | 0.136*** | 0.161*** | 0.093*** | 0.141*** | 0.234*** | 0.274*** |
| | | | | | | | (0.011) | (0.011) | (0.021) | (0.031) | (0.064) | (0.060) |
| Constant | 4.092*** | 4.078*** | 3.985*** | 3.979*** | 3.187*** | 2.725*** | 2.279*** | 1.860*** | 2.207*** | 1.379*** | 3.130*** | 2.772*** |
| | (0.009) | (0.010) | (0.018) | (0.019) | (0.116) | (0.115) | (0.120) | (0.116) | (0.254) | (0.376) | (0.507) | (0.516) |
| Number of observations | 11,463 | 11,463 | 11,463 | 11,463 | 11,463 | 11,463 | 11,463 | 11,463 | 3,301 | 2,085 | 1,213 | 1,130 |
| $R^2$ | 0.010 | 0.019 | 0.029 | 0.038 | 0.105 | 0.127 | 0.189 | 0.229 | 0.160 | 0.248 | 0.217 | 0.318 |
| Fixed Effects | EA | EA | EA | EA | EA | EA | EA | EA | EA | EA | EA | EA |
| Test of difference (GPS and SR; p-values) | 0.0002 | | 0.0001 | | 0.0000 | | 0.0001 | | 0.0298 | | 0.4715 | |

note: *** p<0.01, ** p<0.05, * p<0.1
Robust standard errors in parenthesis, clustered at EA level

**Table 8: Country Level**

| | Malawi | | | | Uganda | | | | Tanzania | | | | Niger | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Per Farm | | Per Acre | | Per Farm | | Per Acre | | Per Farm | | Per Acre | | Per Farm | | Per Acre | |
| | GPS | SR | GPS | SR | GPS | SR | GPS | SR | GPS | SR | GPS | SR | GPS | SR | GPS | SR |
| Log farm size (acres) | 0.372*** | 0.262*** | -0.177*** | -0.241*** | 0.351*** | 0.307*** | -0.154* | -0.169* | 0.256*** | 0.356*** | -0.195** | -0.173** | 0.317*** | 0.087 | -0.313*** | -0.425*** |
| | (0.019) | (0.021) | (0.024) | (0.024) | (0.067) | (0.065) | (0.087) | (0.089) | (0.073) | (0.084) | (0.082) | (0.081) | (0.103) | (0.060) | (0.086) | (0.087) |
| _Plot Characteristics_ | | | | | | | | | | | | | | | | |
| Share of farm with irrigation | 0.241 | 0.178 | 0.291* | 0.209 | 0.003 | 0.076 | 0.092 | 0.185 | 1.453** | 1.391** | 1.274** | 1.216** | 0.751 | 0.838 | 0.514 | 0.617 |
| | (0.163) | (0.166) | (0.160) | (0.169) | (0.302) | (0.334) | (0.279) | (0.285) | (0.662) | (0.648) | (0.548) | (0.565) | (0.944) | (1.042) | (0.871) | (0.943) |
| HH has full or partial tree cover | 0.295*** | 0.293*** | 0.287*** | 0.283*** | 0.151 | 0.136 | 0.125 | 0.124 | 1.618*** | 1.619*** | 1.416*** | 1.394*** | 0.711*** | 0.749** | 0.604** | 0.688** |
| | (0.040) | (0.041) | (0.039) | (0.040) | (0.138) | (0.137) | (0.139) | (0.133) | (0.137) | (0.136) | (0.132) | (0.126) | (0.271) | (0.306) | (0.249) | (0.302) |
| Share of farm with steep slope | 0.034 | 0.035 | 0.014 | 0.025 | -0.244 | -0.404 | -0.198 | -0.276 | 0.040 | -0.051 | -0.080 | -0.234 | 0.142 | 0.140 | -0.046 | -0.008 |
| | (0.084) | (0.083) | (0.088) | (0.083) | (0.284) | (0.313) | (0.282) | (0.309) | (0.363) | (0.380) | (0.323) | (0.328) | (0.405) | (0.395) | (0.318) | (0.314) |
| Share of farm with sandy soil | -0.019 | -0.022 | -0.016 | -0.024 | 0.031 | 0.071 | 0.043 | 0.078 | 0.367** | 0.375** | 0.313** | 0.330** | 0.231 | 0.242 | 0.191 | 0.204 |
| | (0.031) | (0.031) | (0.030) | (0.030) | (0.106) | (0.107) | (0.101) | (0.100) | (0.155) | (0.157) | (0.148) | (0.145) | (0.156) | (0.169) | (0.131) | (0.139) |
| Share of farm with clay soil | 0.001 | -0.009 | -0.003 | -0.009 | 0.040 | 0.107 | 0.081 | 0.078 | 0.290* | 0.310** | 0.197 | 0.212 | 0.459 | 0.434 | 0.357 | 0.353 |
| | (0.028) | (0.029) | (0.028) | (0.028) | (0.161) | (0.157) | (0.153) | (0.153) | (0.149) | (0.152) | (0.140) | (0.144) | (0.315) | (0.311) | (0.267) | (0.264) |
| Share of farm with property rights | -0.022 | -0.008 | -0.001 | 0.006 | -0.152 | -0.151 | -0.216 | -0.167 | 0.309** | 0.288* | 0.238 | 0.204 | 0.073 | 0.164 | 0.007 | 0.130 |
| | (0.108) | (0.112) | (0.110) | (0.109) | (0.152) | (0.150) | (0.149) | (0.146) | (0.157) | (0.158) | (0.149) | (0.144) | (0.277) | (0.287) | (0.242) | (0.258) |
| Share of farm used free but not owned | 0.039 | 0.028 | 0.037 | 0.025 | -0.207 | -0.281 | -0.243 | -0.278 | -0.028 | -0.020 | -0.062 | -0.078 | -0.238 | -0.261 | -0.231 | -0.189 |
| | (0.037) | (0.037) | (0.037) | (0.036) | (0.178) | (0.182) | (0.205) | (0.196) | (0.206) | (0.204) | (0.191) | (0.182) | (0.164) | (0.171) | (0.153) | (0.155) |
| Share of farm rented for a fee | 0.070 | 0.049 | 0.076 | 0.048 | -0.351 | -0.369 | -0.425 | -0.392 | -0.380 | -0.315 | -0.407 | -0.420 | 0.407 | 0.316 | 0.330 | 0.266 |
| | (0.053) | (0.054) | (0.052) | (0.051) | (0.271) | (0.267) | (0.267) | (0.267) | (0.465) | (0.463) | (0.390) | (0.395) | (0.308) | (0.320) | (0.301) | (0.335) |
| _Household Characteristics_ | | | | | | | | | | | | | | | | |
| Log household head age | 0.021 | 0.073** | 0.029 | 0.077*** | -0.020 | 0.019 | -0.047 | -0.003 | 0.356** | 0.347** | 0.204 | 0.217 | -0.476** | -0.416** | -0.384*** | -0.255* |
| | (0.028) | (0.028) | (0.028) | (0.028) | (0.137) | (0.137) | (0.149) | (0.141) | (0.167) | (0.163) | (0.166) | (0.155) | (0.193) | (0.196) | (0.147) | (0.149) |
| Female HH head | -0.125*** | -0.136*** | -0.127*** | -0.134*** | -0.026 | -0.001 | -0.000 | 0.002 | -0.280** | -0.226* | -0.243* | -0.185 | -0.363* | -0.480** | -0.375** | -0.495*** |
| | (0.025) | (0.026) | (0.026) | (0.026) | (0.101) | (0.105) | (0.103) | (0.106) | (0.127) | (0.131) | (0.124) | (0.126) | (0.210) | (0.216) | (0.177) | (0.192) |
| Log household head education (years) | 0.069*** | 0.068*** | 0.072*** | 0.069*** | 0.015 | 0.006 | 0.026 | 0.015 | -0.051 | -0.044 | -0.073 | -0.048 | -0.097 | -0.096 | -0.089 | -0.085 |
| | (0.013) | (0.014) | (0.013) | (0.013) | (0.056) | (0.058) | (0.056) | (0.058) | (0.067) | (0.066) | (0.062) | (0.061) | (0.067) | (0.065) | (0.057) | (0.054) |
| Head's primary occupation is agriculture | 0.045* | 0.061** | 0.043* | 0.064*** | 0.162* | 0.174** | 0.172** | 0.209** | 0.201 | 0.218 | 0.168 | 0.185 | 0.252 | 0.260 | 0.161 | 0.113 |
| | (0.024) | (0.024) | (0.023) | (0.023) | (0.089) | (0.089) | (0.087) | (0.086) | (0.135) | (0.133) | (0.136) | (0.131) | (0.188) | (0.189) | (0.139) | (0.133) |
| Log number of plots on farm | 0.367*** | 0.391*** | 0.373*** | 0.391*** | 0.358*** | 0.307*** | 0.337*** | 0.295*** | 0.077 | 0.072 | 0.003 | 0.028 | 0.169 | 0.232* | 0.154 | 0.216* |
| | (0.034) | (0.036) | (0.033) | (0.034) | (0.084) | (0.084) | (0.082) | (0.082) | (0.134) | (0.137) | (0.125) | (0.127) | (0.123) | (0.131) | (0.106) | (0.114) |
| Log HH population | 0.015 | 0.037* | 0.012 | 0.039** | 0.238** | 0.248** | 0.203** | 0.220** | 0.121 | 0.122 | 0.130 | 0.118 | 0.126 | 0.163 | 0.127 | 0.149 |
| | (0.020) | (0.021) | (0.020) | (0.020) | (0.093) | (0.097) | (0.089) | (0.094) | (0.113) | (0.114) | (0.109) | (0.111) | (0.123) | (0.124) | (0.092) | (0.096) |
| _Inputs_ | | | | | | | | | | | | | | | | |
| Log non-labor inputs value (per acre) | 0.269*** | 0.289*** | 0.270*** | 0.289*** | 0.087** | 0.093*** | 0.113*** | 0.121*** | 0.008 | -0.008 | 0.024 | -0.011 | 0.110* | 0.161*** | 0.082 | 0.144** |
| | (0.013) | (0.013) | (0.013) | (0.013) | (0.034) | (0.034) | (0.043) | (0.042) | (0.035) | (0.036) | (0.046) | (0.048) | (0.058) | (0.056) | (0.080) | (0.071) |
| Log family labor days (per acre) | 0.127*** | 0.149*** | 0.129*** | 0.153*** | 0.285*** | 0.331*** | 0.312*** | 0.341*** | 0.209*** | 0.184*** | 0.187*** | 0.168*** | 0.171** | 0.234*** | 0.167** | 0.227*** |
| | (0.016) | (0.017) | (0.016) | (0.016) | (0.079) | (0.080) | (0.088) | (0.085) | (0.071) | (0.070) | (0.069) | (0.065) | (0.075) | (0.067) | (0.067) | (0.067) |
| Log hired labor days (per acre) | 0.132*** | 0.149*** | 0.137*** | 0.165*** | 0.127*** | 0.140*** | 0.162*** | 0.153*** | 0.129*** | 0.118*** | 0.102* | 0.120** | 0.102*** | 0.115*** | 0.193*** | 0.173** |
| | (0.010) | (0.010) | (0.011) | (0.012) | (0.026) | (0.027) | (0.035) | (0.033) | (0.039) | (0.038) | (0.054) | (0.057) | (0.038) | (0.038) | (0.069) | (0.080) |
| Constant | 2.465*** | 2.057*** | 2.424*** | 2.014*** | 2.429*** | 2.050*** | 2.369*** | 2.041*** | -1.470** | -1.424** | -0.566 | -0.599 | 3.676*** | 3.338*** | 3.460*** | 2.923*** |
| | (0.122) | (0.121) | (0.122) | (0.120) | (0.634) | (0.630) | (0.674) | (0.641) | (0.733) | (0.722) | (0.696) | (0.659) | (0.705) | (0.684) | (0.514) | (0.522) |
| Number of observations | 8,799 | 8,799 | 8,799 | 8,799 | 831 | 831 | 831 | 831 | 1,202 | 1,202 | 1,202 | 1,202 | 631 | 631 | 631 | 631 |
| $R^2$ | 0.418 | 0.403 | 0.205 | 0.245 | 0.488 | 0.479 | 0.259 | 0.285 | 0.440 | 0.446 | 0.286 | 0.271 | 0.259 | 0.235 | 0.217 | 0.366 |
| Fixed Effects | EA | EA | EA | EA | EA | EA | EA | EA | EA | EA | EA | EA | EA | EA | EA | EA |
| Test of difference (GPS and SR; p-values) | 0.0000 | | 0.0005 | | 0.3754 | | 0.6972 | | 0.1539 | | 0.6660 | | 0.0169 | | 0.0447 | |

note: *** p<0.01, ** p<0.05, * p<0.1

Robust standard errors in parenthesis, clustered at EA level

**Table A1: Probability of GPS Measurement**
Dependant Variable: 1 if GPS is measured
Marginal Effects

| | Malawi | Uganda | Tanzania | Niger |
|---|---|---|---|---|
| *Plot Characteristics* | | | | |
| SR plot size (acres) | -0.0001 | -0.010*** | 0.003 | -0.0005 |
| | [0.000] | [0.001] | [0.420] | [0.807] |
| Steep slope | -0.007 | -0.072 | -0.125*** | -0.151*** |
| | [0.407] | [0.161] | [0.000] | [0.000] |
| Number of plots per HH | 0.0006 | -0.012** | -0.002 | -0.018** |
| | [0.719 | [0.045] | [0.647] | [0.015] |
| Property Rights | -0.03 | 0.043 | 0.050** | -0.045 |
| | [0.123] | [0.264] | [0.043] | [0.308] |
| Rented In | -0.025*** | -0.313*** | -0.202*** | -0.074* |
| | [0.000] | [0.000] | [0.000] | [0.086] |
| Used Free | -0.007 | -0.260*** | -0.174*** | -0.124*** |
| | [0.182] | [0.000] | [0.000] | [0.000] |
| Distance from HH (km) | - | - | -0.028*** | -0.072*** |
| | | | [0.000] | [0.000] |
| Distance from HH (mins): | | | | |
| 15 - 30 mins | - | -0.317*** | - | - |
| | | [0.000] | | |
| 30 -60 mins | - | -0.531*** | - | - |
| | | [0.000] | | |
| 1 - 2 hours | - | -0.617*** | - | - |
| | | [0.000] | | |
| More than 2 hours | - | -0.626*** | - | - |
| | | [0.000] | | |
| *Household Characteristics* | | | | |
| Head age | 0.0004*** | 0.001 | -0.0001 | -0.003*** |
| | [0.000] | [0.366] | [0.928] | [0.003] |
| Female HH Head | 0.010** | 0.042 | 0.01 | -0.060 |
| | [0.016] | [0.112] | [0.620] | [0.215] |
| Head years of education | 0.0003 | -0.002 | -0.003 | 0.006 |
| | [0.542] | [0.476] | [0.191] | [0.269] |
| HH population | -0.002** | -0.007* | -0.005* | 0.001 |
| | [0.015] | [0.057] | [0.066] | [0.680] |
| Urban | -0.092*** | -0.329*** | 0.031 | 0.050 |
| | [0.000] | [0.000] | [0.294] | [0.383] |
| N° | 18426 | 5003 | 4797 | 5609 |

P-values in brackets
* Significant at 10%, **Significant at 5%, ***Significant at 1%
°Limited to cultivated plots with SR areas