

Could the Debate Be Over?

Errors in Farmer-Reported Production and Their Implications for the Inverse Scale-Productivity Relationship in Uganda

Sydney Gourlay

Talip Kilic

David Lobell



WORLD BANK GROUP

Development Economics
Development Data Group
September 2017

Abstract

Based on a two-round household panel survey conducted in Eastern Uganda, this study shows that the analysis of the inverse scale-productivity relationship is highly sensitive to how plot-level maize production, hence yield (production divided by GPS-based plot area), is measured. Although farmer-reported production-based plot-level maize yield regressions consistently lend support to the inverse scale-productivity relationship, the comparable regressions estimated with maize yields based on sub-plot crop cutting, full-plot crop cutting, and remote sensing point toward constant returns to scale, at the mean as well as throughout the distributions of objective measures of maize yield. In deriving the much-debated coefficient for GPS-based plot area, the maize yield regressions control for objective measures of soil fertility, maize genetic

heterogeneity, and edge effects at the plot level; a rich set of plot, household, and plot manager attributes; as well as time-invariant household- and parcel-level unobserved heterogeneity in select specifications that exploit the panel nature of the data. The core finding is driven by persistent overestimation of farmer-reported maize production and yield vis-à-vis their crop cutting-based counterparts, particularly in the lower half of the plot area distribution. Although the results contribute to a larger, and renewed, body of literature questioning the inverse scale-productivity relationship based on omitted explanatory variables or alternative formulations of the agricultural productivity measure, the paper is among the first documenting how the inverse relationship could be a statistical artifact, driven by errors in farmer-reported survey data on crop production.

This paper is a product of the Development Data Group, Development Economics. 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 sgourlay@worldbank.org.

The Policy Research Working Paper Series disseminates the findings of work in progress to encourage the exchange of ideas about development issues. An objective of the series is to get the findings out quickly, even if the presentations are less than fully polished. The papers carry the names of the authors and should be cited accordingly. The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors. They do not necessarily represent the views of the International Bank for Reconstruction and Development/World Bank and its affiliated organizations, or those of the Executive Directors of the World Bank or the governments they represent.

Could the Debate Be Over?

Errors in Farmer-Reported Production and Their Implications for the Inverse Scale-Productivity Relationship in Uganda

Sydney Gourlay[‡], Talip Kilic[†], and David Lobell^{#1}

JEL Codes: C23, C83, Q12.

Keywords: Maize, Yield Measurement, Plot Area Measurement, Inverse Scale-Productivity Relationship, Crop Cutting, Remote Sensing, Household Surveys, Uganda, Sub-Saharan Africa.

¹ The senior authorship is shared between Sydney Gourlay and Talip Kilic. [‡] Corresponding author. Ph.D. Candidate, Department of Economics, The American University, Washington, DC, and Survey Specialist, Living Standards Measurement Study (LSMS), Survey Unit, Development Data Group, World Bank. sgourlay@worldbank.org [†] Senior Economist, Living Standards Measurement Study (LSMS), Survey Unit, Development Data Group, World Bank, Rome, Italy. tkilic@worldbank.org [#] Professor, Center on Food Security and the Environment, Stanford University, Stanford, CA. dlobell@stanford.edu. The authors would like to thank for their comments (in alphabetical order) Leigh Anderson, Leah Bevis, Sam Desiere, Nancy McCarthy, the participants of (i) the USAID Uganda Agriculture Market Systems Workshop (March 2017, Kampala, Uganda), (ii) the World Bank Land and Poverty Conference (March 2017, Washington, DC), (iii) the Centre for the Study of African Economies (CSAE) conference (March 2017, Oxford, UK); (iv) the FAO Agricultural Development Economics Seminar and the IFAD Research and Impact Assessment Seminar (April 2017, Rome, Italy), (v) the University of Copenhagen Development Economics Seminar (May 2017, Copenhagen, Denmark), and (vi) the XV EAAE Congress (August 2017, Parma, Italy).

1. Introduction

Worldwide, 475 million farms, constituting 84 percent of a total of 570 million farms, are estimated to be less than 2 hectares (Lowder et al., 2016), and (smallholder) agricultural activities constitute an integral part of livelihoods in rural areas that are home to nearly 70 percent of the population in low-income countries.² In Africa specifically, the average share of rural household income stemming from agriculture could be up to 69 percent (Davis et al., 2017), and research has consistently documented higher rates of expected poverty reduction associated with agricultural vis-à-vis nonagricultural growth (see Dorosh and Thurlow, 2016 and the studies cited therein).

The importance of agriculture for development is recognized in the formulation of the Sustainable Development Goal (SDG) Targets 2.3 and 2.4, which require doubling of agricultural productivity and incomes of small-scale food producers, and ensuring sustainable food production systems and implementing resilient agricultural practices that increase productivity and production. Both targets are associated with indicators³ that rely on crop production and land area information sourced from household or farm surveys, and the documented effects of measurement on accurate measurement and analysis of land productivity underscore the need for high-quality survey data for cross-country monitoring, broader economic research, and policy formulation (Carletto et al., 2013, 2015; Kilic et al. 2017a; 2017b).

An often observed yet insufficiently explained puzzle in smallholder agriculture is that of the inverse relationship between scale (in terms of farm or plot size) and (land) productivity (henceforth referred to as the IR). The existence of the IR could have a direct bearing on policy formulation as it, for instance, relates to reforms targeting land and non-land input markets; land redistribution programs; and identification of beneficiary target universe for agricultural development programs. If the IR does indeed exist, and is not an artifact of measurement error in the data, that would suggest governments might encourage redistribution of land, favoring smaller allotments – even if the operational implications of such recommendation are not clear. Alternatively, the existence of the IR could simply temper concerns around plots becoming smaller or, more broadly, “small” plots, often managed by females and poorer farmers (Kilic et al., 2015a; 2015b).

The body of literature on the IR is rich, and yet inconclusive. In the African context, Larson et al. (2014) revisit the IR debate with a focus on maize, using data from a wide array of household and farm surveys conducted from 1999 to 2009 across the continent, and provide support to the existence of the IR both at farm- and plot-levels. Following Eastwood et al. (2010), it has been suggested that the IR may be explained by (i) lower supervision costs on smaller holdings and

² The statistic is for 2016, the most recent year available through <http://data.worldbank.org>.

³ The final list of SDG indicators can be found in Inter-Agency and Expert Group on Sustainable Development Goal Indicators (2016).

plots (Feder, 1985); (ii) missing or incomplete factor markets (Barrett, 1996; Eswaran and Kotwal, 1986); (iii) omitted variables, in particular, controls for farmer ability and land quality (Assuncao and Braido, 2007; Assuncao and Ghatak, 2007; Benjamin, 1995; Bhalla and Roy, 1988; Lamb, 2003), and (iv) errors in land area measurement (Lamb, 2003).

Working backwards, on hypothesis (iv), Carletto et al. (2013) in a Uganda-specific study, and Carletto et al. (2015) in a cross-country study focused on Sub-Saharan Africa, document that land areas, on average, are over-reported by farmers, but that the IR persists even after objective, GPS-based land area measures are used in empirical analyses.⁴ On hypothesis (iii), Nkonya et al. (2004) and Barrett et al. (2010) show that controlling for objectively-measured soil quality based on laboratory analyses does not explain the IR at the plot-level. On hypothesis (ii), Ali and Deininger (2015) show that at the farm-level, the IR can be explained by using a productivity measure that nets out the labor input valued at market wages.⁵ A more recent hypothesis not reviewed by Eastwood et al. (2010) suggests that land productivity may be greater along the edge of plots as labor may be concentrated in those visible areas (Bevis and Barrett, 2017). Smaller plots have a greater ratio of visible edge area to less-visible interior area, which if edges are indeed more productive, may explain the often-observed IR. Bevis and Barrett (2017) provide support for this hypothesis in the context of Uganda.

In debating the IR, however, the question of potential measurement error in production figures has rarely been examined. This is a profound oversight considering that the studies quoted above all measure agricultural productivity based on farmer-reported production. This is likely attributable (at least in part) to the fact that data on objectively-measured agricultural production via crop cutting is rarely available in large-scale household survey operations due to resource constraints. Yet, severe systematic biases have been found in other farmer-estimated data, such as plot area (see Carletto et al., Forthcoming, and the studies cited therein). Farmer-reported production estimates are exceptionally complicated. The use of non-standard production units, various conditions and states of crop harvests, and the reporting of permanent crop harvests, for starters, threaten the quality of farmer-estimated production. Additionally, humans often exhibit an inclination to round off numbers, as shown for land area measurement in Carletto et al. (2015), which may bias production estimates.

Using unique panel data from a methodological study on maize productivity in Eastern Uganda that includes self-reported, as well as highly-supervised crop-cutting and remotely-sensed production and yield estimates, this study aims to determine whether the IR can be explained by measurement error in self-reported production estimates. The yield measures, irrespective of the approach to production measurement, are anchored in GPS-based plot area measurement. Overall,

⁴ Kilic et al. (2017b) demonstrate that IR persists at the plot-level in the context of Tanzania and Uganda even after accounting for selection bias in missing GPS-based plot area measurements through the use of multiple imputation.

⁵ Desiere and Jolliffe (2017) aptly argue that the missing/incomplete factor market hypothesis apply strictly to exploration of IR at the farm-level, as opposed to within-farm plot-level.

we provide unambiguous support for the sensitivity of the plot-level IR to the choice of the method by which maize production and yield are computed. While farmer-reported production-based maize yield regressions consistently imply diminishing returns to GPS-based plot area, the comparable regressions estimated with maize yields based on sub-plot crop cutting, full-plot crop cutting, and high-resolution satellite imagery-based remote sensing point towards constant returns to scale.

In view of the aforementioned hypotheses put forth for the IR, it is important to note that the results are robust to the inclusion of objective measures of soil fertility, maize genetic heterogeneity and edge effects at the plot-level; a rich set of plot, household and plot manager attributes; as well as household and parcel fixed effects in select specifications that exploit the panel nature of the data. In other words, irrespective of the exhaustive list of controls and panel estimation, the IR exists while using farmer-reported maize production, and ceases to do so when maize yield is anchored in objective measurement methods. In fact, the IR persists throughout the distribution of plot-level maize yields that are based on farmer-reported production, while constant returns to scale prevails across the productivity distribution when one uses objective yield measures.

Our core finding is driven by persistent over-estimation of farmer-reported maize production and yield vis-à-vis their crop cutting-based counterparts, particularly in the lower half of the plot area distribution. Though our findings contribute to a larger (and renewed) body of literature questioning the inverse scale-productivity relationship based on omitted explanatory variables or alternative formulations of the agricultural productivity measures, the analysis, together with Desiere and Jolliffe (2017), is the first documenting how the IR could be a statistical artifact, driven by errors in self-reported survey data on crop production.

The paper is organized as follows. Section 2 discusses the conceptual basis for errors in farmer-reported crop production and their potential implications for the IR. Section 3 provides relevant background on the Ugandan context. Section 4 describes the data. Section 5 lays out the empirical strategy. Section 6 presents the results. Section 7 concludes.

2. Conceptual Basis for Potential Errors in Farmer-Reported Crop Production

The quality of farmer-reported estimates of production is degraded by a myriad of issues, summarized in Table 1. Human performance and behavior are subject to a number of constraints, one of them being finite memory. Recall bias can, therefore, pose a serious threat to the quality of farmer-reported estimates of production. This is especially relevant if there is partial green harvest (for immediate consumption or sale) or if permanent or extended-harvest crops, such as cassava,

are grown.⁶ In these situations, aggregating all of the production from the previous completed agricultural season becomes a taxing mental exercise.

Also, inherent in human behavior is the inclination to round off values. For instance, recent research has illustrated the persistence of rounding in plot area estimates and the ensuing effects on data quality (Carletto et al., 2015; Forthcoming). The same problem is likely present in the estimation of production. Rounding of production estimates may pose a bigger problem on plots with lower production, such as smaller plots, as the rounding error relative to true production is likely greater on these plots. Suppose, for the sake of example, that farmer A, with true production of 1.5 bags, reports that he produced 2 bags. Farmer B has a true production of 11.5 bags, but he reports 12 bags of production. Farmer A overstates his production by 33.3 percent, while farmer B only overstates by 4.3 percent.

Further, depending on the context and farmer characteristics, intentional bias may be at play. Farmers may be inclined to understate production if they perceive an incentive to do so. Perceived incentives may come in the form of eligibility for various assistance programs or a possible threat of taxation, for example. Social desirability bias may lead farmers to overstate their production in order to appear successful.

Setting the challenges with human behavior aside, aggregating farmer-reported production data for analysis poses several challenges, as summarized by Oseni et al. (2017). First, farmers frequently utilize non-standard measurement units for the quantification of production. These units vary across countries, and often exhibit within-country variation across space and time. Even if the questionnaire instrument permits the use of non-standard measurement units for recording production quantities, for standard analyses, these values must either be monetized or converted into kilogram (kg) equivalent terms. In the case of monetization, if the common measurement units used for production quantification differ from the common units associated with crop sales, the valuation of production will be a challenge without converting crop production and sales into kg-equivalent terms. In the case of kg-equivalent conversion, one would need to rely on crop-unit-specific conversion factors (possibly differentiated by region), which are often either unavailable to the researcher or available but not documented adequately and/or suspected to be out of date/in need of further validation.

A factor that further mediates the success of the kg-equivalent conversion is the issue of crop condition and state, and in the case of cereals, the ability to express production in kg-grain-equivalent terms. Put differently, a farmer that reports two 100 kg bags of green maize that is on the cob would have drastically different production than a farmer that reports two 100 kg bags of maize in dried grain form. If the questionnaire instrument does not allow for the distinction of crop condition and state, the researcher would need to make assumptions regarding the crop conditions

⁶ Extended-harvest refers to harvest on an ongoing basis.

and states associated with the reported production quantities. Even if the questionnaire instrument allows for the distinction of crop condition and state, the conversion factor database would need to cover each crop-unit-condition-state combination – a more demanding challenge for survey implementers to address compared to having a simpler conversion factor database at the crop-unit-level.

As the literature on the IR largely overlooks the abovementioned issues with relying on farmer-reported production values, we now turn to the description of the Ugandan context and the data that allow us to explore whether, and the extent to which, errors in farmer-reported production information exist and how, if at all, they affect the often-observed IR.

3. Country Context

Uganda is a landlocked East African nation of approximately 39 million people, with an annual population growth rate of 3.25 percent and 83.9 percent of the population living in rural areas (2015). National and rural rates of the population living in poverty, with respect to the national poverty line, are estimated at 19.5 and 22.4 percent, respectively (2012), and gross domestic product per capita in current US dollars stands at 705 (2015). The economy is heavily dependent on agriculture, such that agricultural land constitutes 71.9 percent of total land area (2014), agriculture value added corresponds to 25.8 percent of the GDP (2015), and agricultural employment makes up 71.7 percent of total employment (2013).⁷ Agriculture, specifically increasing agricultural production and productivity, is identified as an essential vehicle for wealth creation in the country's key policy documents (GoU, 2013; 2015; MAAIF, 2013).

Maize is a major staple, commercial, and export crop in Uganda. It is the leading cereal crop grown in almost all parts of the country. In Eastern Uganda, the country's leading maize producing region (UBOS, 2010), the crop accounts for the highest share (25 percent) of crop income (World Bank, 2016). At the same time, Eastern Uganda, following Northern Uganda, is also the region with the highest concentration of the country's poor, and the latest estimate of the regional absolute poverty rate stands at 24.5 percent (World Bank, 2016). Commercialization of maize is relatively low in Eastern Uganda, with less than 41 percent of maize producing households selling *any* quantity of maize.⁸

According to FAOSTAT data depicted in Figure 1 and Figure 2, the trends in maize area harvested and maize yield are positive for the period of 1995-2014. The national area harvested increased steadily over time, from 571,000 hectares to 1.1 million hectares in 2014. The national yield

⁷ The official World Bank statistics on Uganda are obtained from <https://goo.gl/IqGLQh>. The reference years are noted in parentheses.

⁸ The statistic is based on the Uganda National Panel Survey (UNPS) 2015/16 data.

fluctuated around 1.5 tons per hectare mark from 1995 to 2007, but has been above 2.3 tons per hectare in the period of 2008-2014. The latest national yield estimate in 2014 was 2.5 tons per hectare. Recent FAOSTAT yield estimates, however, are markedly higher than those computed from household survey data, including from the methodological experiment informing this study, which employs objective measurement methods. These discrepancies in yield estimates call into question, at least for Uganda, the validity of FAOSTAT figures.

4. Data

4.1. Overview

MAPS: Methodological Experiment on Measuring Maize Productivity, Soil Fertility and Variety is a two-round household panel survey that was conducted in Eastern Uganda to test the relative accuracy of subjective approaches to data collection vis-à-vis objective survey methods for maize yield measurement, soil fertility assessment, and maize variety identification. The survey has been implemented by the Uganda Bureau of Statistics, with technical and financial assistance provided by an inter-agency partnership that is led by the World Bank Living Standards Measurement Study (LSMS), using the Survey Solutions Computer Assisted-Personal Interviewing (CAPI) platform.⁹

4.2. Sampling Design

In Round I, the MAPS fieldwork was conducted during the first rainy season of 2015, from April to October 2015, in Eastern Uganda, the top maize-producing region of the country. The sample was composed of 75 enumeration areas (EAs) that were selected from the 2014 Population and Household Census (PHC) EA frame and that were distributed across 3 strata, namely (1) Sironko district (15 EAs), (2) Serere district (15 EAs), and (3) a 400 km² remote sensing tasking area spanning Iganga and Mayuge districts (45 EAs). In each stratum, the EAs were sampled with probability proportional to size, in accordance with the pre-dissemination 2014 PHC EA-level household counts.

⁹ The technical assistance to MAPS I (2015) and MAPS II (2016) was provided through the World Bank LSMS – Minding the (Agricultural) Data Gap Research Program, funded by UK Aid. In MAPS I, MAPS implementation was financed by the World Bank LSMS – Minding the (Agricultural) Data Gap Research Program, the Global Strategy to Improve Agricultural and Rural Statistics, led by the Food and Agriculture Organization of the United Nations (FAO), and the CGIAR Standing Panel on Impact Assessment. In MAPS II, MAPS implementation was financed by the World Bank Innovations in Big Data Analytics Program, the World Bank Trust Fund for Statistical Capacity Building – Innovations in Development Data Window, and the CGIAR Standing Panel on Impact Assessment. In MAPS I and MAPS II, Terra Bella provided free high-resolution satellite imagery for the MAPS remote sensing tasking area for research purposes. The technical partners included (i) the World Agroforestry Center on soil fertility measurement, (ii) the CGIAR Standing Panel on Impact Assessment on maize variety identification, and (iii) Stanford University Center on Food Security and Environment on remote sensing.

In each sampled EA, a household listing exercise was conducted to identify, separately, the list of households that were cultivating at least 1 pure stand maize plot, and the list of households that were cultivating at least 1 intercropped maize plot, on which maize is self-identified to be the dominant crop. Given the interest in the validation of survey methods in key sub-samples and the fact that approximately two-thirds of all maize plots are intercropped in Uganda, the original intention had been to select, at random, 6 households from each of the pure stand and intercropped universes of households of an EA, and ensure an even sample split by maize cultivation status.

Still, due to the low incidence of pure stand households, and the cases in which pure stand households would switch to intercropping status between the household listing and the first interview, the sample at the start of MAPS I fieldwork was composed of 900 households, of which 385 were pure stand (43 percent) and 515 were intercropped (57 percent). Within the remote sensing tasking area specifically, the MAPS I fieldwork started out with 540 households, of which 249 were pure stand (46 percent) and 291 (54 percent) were intercropped. In each MAPS household, 1 maize plot, matching the household cultivation status, was selected at random by the Survey Solutions CAPI application for crop cutting (for objective yield measurement) and soil sampling (for objective soil fertility analysis). The variety identification component was implemented among the 540 MAPS I households residing in the remote sensing tasking area (specific to the plots on which crop cutting and soil sampling took place), as explained below.

In MAPS II, the MAPS fieldwork was conducted during the first rainy season of 2016, from June to October 2016. The field teams attempted to track and re-interview 540 households that had been interviewed in MAPS I in the 400 km² remote sensing tasking area cutting across Iganga and Mayuge districts. Appendix I provides the MAPS II household tracking protocol that was followed by the field teams. The MAPS II fieldwork successfully interviewed 489 out of 540 households, and the sample informing our analyses is composed of 440 households for which we have obtained crop cutting measures in both rounds.¹⁰ Further, as in Round I, 1 maize plot was selected from each household for crop cutting and variety identification components, in accordance with the following rules. Whenever possible, a plot was selected among those that were matching the household cultivation status in Round I. Preference was also given such that a plot would be selected from the same parcel that had contained the plot selected in Round I. If multiple plots match the household cultivation status, the CAPI application selected one plot at random. Appendix II lays out the MAPS II plot selection protocol that was implemented by the field teams.

¹⁰ 34 out of 51 households that we did not interview in MAPS II were due to the fact that they were not cultivating maize in the first season of 2016. The remaining 17 households can be broken down as follows: 5 households could not be tracked or were outside of the tracking area defined as the Iganga and Mayuge districts (5); 4 households had suffered total crop loss prior to post-planting interview; 7 households had already harvested their maize by the post-planting interview; and 1 household refused. The final analysis sample of 440 households vis-à-vis the original 100 records that had at least been subject to the post-planting interview in MAPS I do not exhibit statistically significant differences in terms of their MAPS I yield measures and control variables for our regressions. The mean comparisons are available upon request. In line with this finding, our multivariate regressions are robust to the use of inverse predicted response probabilities in MAPS I as attrition weights.

4.3. Fieldwork

In each round, there were three visits to each household, namely post-planting, crop-cutting, and post-harvest. During the post-planting visit, each household was administered a farm survey that collected information on (1) age, sex, education, economic activities, ethnicity, religion, and extension service receipts for all household members; (2) household dwelling and ownership of consumer durables and farm assets; (3) area, tenure, and individual-disaggregated ownership and rights for all parcels that were owned and/or cultivated by households during the reference rainy season; and (4) information on area, cultivation pattern, management and decision-making, conservation agricultural activities, farmer-assessed soil attributes and quality, pre-harvest labor and seed inputs for all maize plots that were cultivated during the reference rainy season.¹¹ The plot-level information was intended to be solicited from the corresponding plot manager. Specific to our analysis sample, the rate at which the plot-level information was solicited from the intended plot manager stands at 83 percent in both MAPS I and MAPS II. Following the completion of the household post-planting interview, the enumerator visited the randomly selected maize plot, measured its area and saved its boundaries on a Garmin eTrex 30 handheld GPS device, and set up crop cut sub-plots, in accordance with the international best practices, for later harvesting and weighing. A local crop monitor was recruited in each EA to ensure that the designated crop cut sub-plots were not harvested by the households until the subsequent visit. Specific to MAPS I, top and sub-soil samples were also obtained by the enumerators during the post-planting period, for objective soil fertility testing, as detailed below.

During the crop cutting visit, the enumerator harvested the crop cut sub-plots in order to obtain objectively measured harvest quantities. Two household members and the local crop cut monitor helped in the harvesting and shelling of the crop cut harvests that were weighed initially by the enumerator using an industrial digital scale. The samples were later transferred to a centralized facility for additional drying, moisture measurement, and final weighing. At crop cutting, the manager of the randomly selected maize plot also provided information on the morphological attributes of the maize plants on his/her plot with the help of a photo aid, as described below. In the final, post-harvest, visit, farmer-reported information on total plot-specific maize production, non-labor inputs and harvest labor inputs was solicited for all maize plots that were cultivated during the reference season. The post-harvest visit was scheduled within a 2-month period following the completion of each household's harvest.

¹¹ A parcel is conceptualized as a continuous piece of land under a common tenure system, while a plot is defined as a continuous piece of land on which a unique crop or a mixture of crops is grown, under a uniform, consistent crop management system, not split by a path of more than one meter in width, and with boundaries defined in accordance with the crops grown and the operator. Therefore, a parcel can be made up of one or more plots. This distinction is key since for the purposes of within-farm analysis of agricultural productivity, the ideal is to capture within-parcel, plot area measurements linked with plot-level measurement of agricultural production

4.4. Key Measurement Domains and Methods

The measurement methods in each of the key measurement domains of interest, namely maize production, plot area, and soil fertility, are summarized below.

4.4.1. Plot Area Measurement

After walking the perimeter of a given plot with the plot manager to identify the boundaries, the enumerators re-paced the perimeter and measured the area with a Garmin eTrex 30 handheld GPS device. The area was recorded on the questionnaire in square meters, and the raw GPS track outline was stored for the remote sensing work program; linking relevant geospatial variables to the plot location; and calculating plot shape metrics.¹² As noted above, the competing yield measures in our study are all anchored in GPS-based plot area measurement.

4.4.2. Maize Production Measurement

4.4.2.1. Crop Cutting

Crop cutting has been recognized as the gold standard for yield measurement since the 1950s by the Food and Agriculture Organization of the United Nations (FAO). Besides the cost- and supervision-intensive nature of the exercise, several concerns have been raised regarding the accuracy of the method. Even if one places only one random crop cutting sub-plot within the sampled plot, the resulting yield estimate may carry a sampling error if the yields exhibit within-plot heterogeneity. (1) More thorough-harvesting of crop cut sub-plots vis-à-vis the typical farmer harvesting practices, (2) possible rounding of crop cut production estimates obtained through scales; (3) using faulty or inappropriate scales; (4) omitting to net out the weight of the measurement container from the measured production; (5) including plants that fall outside of the sub-plot; and (6) non-random placement of crop cut sub-plots have also been suggested as possible sources of error (Fermont and Benson, 2011).

By implementing a well-supervised crop cutting operation that relied on (a) random sub-plot placement and high-precision digital weighing scales in both rounds, (b) full plot harvests in a subsample of MAPS II plots, and (c) field staff and local crop cut monitors that were intensively trained to circumvent the abovementioned criticisms, our crop cutting-based maize yield estimates, in comparison to those that rely on farmer-reported production, are assumed, and later shown, to be more accurate approximations of the true levels. Further, given our interest in the IR, the main

¹² Each plot manager was also asked to estimate the full area in acres, consistent with the practice of the UNPS. Up to two decimal points were recorded. However, we use the GPS-based plot areas for all analyses, given the systematic errors in farmer-reported plot areas that have been documented, as reviewed above, in Uganda and elsewhere in Sub-Saharan Africa.

assumption in the estimation of our production functions is that the error in crop cutting estimates is independent of the plot size. Given the random placement of the crop cut sub-plots whose size did not vary by plot size, this assumption should be tenable, as also maintained by Desiere and Jolliffe (2017).

Appendix III provides the MAPS sub-plot and full plot crop cutting protocol in detail. In Round I, a 4x4 meter subplot (divided into four 2x2m quadrants) and a separate 2x2 meter subplot were laid on the chosen maize plot during the post-planting visit following a strict protocol to ensure the location of the subplots was random, as described in Appendix III. The subplots were cordoned off until harvest, and were supervised by the local crop cut monitors between the post-planting and the crop cutting visits. Each plot manager was asked not to harvest any crop from the sub-plots until the crop cutting visits, and not to manage the sub-plot any differently than the rest of the plot. These messages, first communicated by the enumerator, were intended to be enforced by the local crop cut monitors.

During the crop cutting visit, the shelled maize harvests tied to each of the five 2x2m quadrants were weighed and barcoded separately in the field, and were reweighed at a central location in Kampala under strict supervision following additional drying (once the moisture content was in the range of 12 to 14 percent). At the time of the final weighing, the moisture content of each sample was captured as to standardize all crop cut sample weights used for our analyses at 12 percent moisture. The MAPS I sub-plot crop cutting based plot-level maize production estimates are computed by multiplying the combined crop cut sub-plot production across the 20m² area covered by the combination of the 4x4m and the 2x2m subplots by the ratio of the entire GPS-based plot area in m² to 20m².¹³

In MAPS II, only one 8x8 meter sub-plot (divided into four 4x4m quadrants) was laid on each plot.¹⁴ The harvests tied to each of the four 4x4m quadrants were weighed and barcoded separately in the field, and the rest of the protocols for crop cut sub-plot supervision and harvest management were identical to those followed in MAPS II.¹⁵ The MAPS II sub-plot crop cutting based plot-level

¹³ While not reported here, the MAPS I average sub-plot crop cutting plot-level maize yield was not sensitive whether one used the 20m² crop cut sub-plot area across both sub-plots; the 16m² crop cut sub-plot area covered by the 4x4m sub-plot; or the 4m² crop cut sub-plot area covered by the 2x2m subplot. These results are available upon request.

¹⁴ The change to the number and size of the sub-plots in MAPS II was underlined by two factors. First, in MAPS I, within the remote sensing tasking area, the difference between the average maize based on the 4x4m subplot and the comparable statistic based on the 2x2m subplot was not statistically significant. Second, the standard deviation among the yields obtained from 5 2x2 crop cut sub-plots was used to calculate the standard error of the mean yield for the field, assuming that the variability within the sample was representative of the variability throughout the field. A large fraction of the fields had standard errors above 10, even 20, percent of the mean, especially for fields with relatively low estimates of mean yield. This in turn raised a question around the extent to which, say, the yield based on a 4x4m sub-plot could be deemed representative of the yield based on the entire plot area. Taking these observations and other logistical considerations into account, the crop cut sub-plot area was increased to 8x8m in MAPS II, with each of the 4x4m quadrant harvests tracked separately to ensure maximum comparability to MAPS I.

¹⁵ In cases where the respondent harvested, despite the instructions during the post-planting visit, a portion of the crop-cutting sub-plot prior to the crop cutting visit, the crop-cutting production figure was inflated by the percentage of the

maize production estimates are computed by multiplying the crop cut sub-plot production across the 64m² area covered by the 8x8m subplot by the ratio of the entire GPS-based plot area in m² to 64m².¹⁶

In addition, prior to the start of the MAPS II fieldwork, half of the target household population was chosen at random, within each of the pure stand and intercropped domain in each EA, to be subject to a full-plot crop cut, as part of which the entire area of the selected plot was harvested, shelled and weighed by the enumerator, with help from the crop cut assistants recruited from the household members and the EA-specific crop cut monitor. On the plots selected for full-plot harvest, the harvest of the designated 8x8m subplot was weighed separately from the full-plot harvest to allow for comparative yield analysis. The full-plot harvests were only weighed in the EAs as their transport to and additional drying and reweighing at a central location was deemed logistically infeasible.

4.4.2.2. Remote Sensing

There is a longstanding and active literature on using remote sensing for agricultural monitoring and assessment. Among more recent advances, the GEOGLAM effort to provide in-season forecasting for major cropping regions (Becker-Reshef et al. 2010, Franch et al. 2015) represents an important step forward, as it has operationalized yield forecasting at the national scale. However, because it is based primarily on high temporal, coarse spatial resolution (~5km) data, it is of little value for field-level assessment. At the same time, many studies have evaluated correlations between different reflectance indices and field or sub-field scale crop biomass or yield in specific site/years (e.g., Shanahan et al. 2001, Lobell et al. 2003, Sibley et al. 2014). Although these studies have demonstrated the potential for accurate satellite-based yield estimates, they generally (i) are for large-scale commercial systems and (ii) rely on calibrations that are specific to sites and/or image timings and thus do not generalize well across broad regions. A recent effort by Lobell et al. (2015) has developed a more general approach to combining crop simulation models and remote sensing data to map yields at the field scale, but this approach so far has been tested at the field scale only in large, homogeneous fields of the U.S.

sub-plot pre-harvested, as reported by the enumerator, based on his/her observation, on the crop-cutting questionnaire. In establishing the pre-harvest percentage, the enumerators were trained extensively to distinguish pre-harvest from crop damage. In cases where the full sub-plot was pre-harvested by the farmer, the observation was dropped from analysis. Of the remaining 440 observations in each wave, 8.4 percent and 20 percent of households pre-harvested a portion of the sub-plot in MAPS I and II, respectively. Of those plots on which pre-harvesting took place, the mean share of the sub-plot pre-harvested was 28.5 percent in MAPS I and 26.2 percent in MAPS II. The crop cut yields were not adjusted for any crop damage that may have materialized between the post-planting and the crop cutting visits.

¹⁶ While not reported here, the MAPS II average sub-plot crop cutting plot-level maize yield was not sensitive whether one used the 64m² crop cut sub-plot area or the 16m² crop cut sub-plot area covered by a randomly selected 4x4m quadrant within the 8x8m sub-plot. These results are available upon request.

Several reviews of the sector point to the key needs of (i) better integration of different data sources, including the new higher resolution data available from several providers, (ii) development of more robust algorithms that can be applied in many different settings without additional calibration, and (iii) more complete and accurate data sets of ground-based measures of crop productivity in order to rigorously establish the accuracy of remote sensing approaches (Gallego et al., 2010; Atzberger, 2013; Lobell; 2013). The MAPS remote sensing work program addresses all of these needs in a unique way. To our knowledge, no other research groups working on agriculture have comparable access to high resolution imagery from the private sector; approaches that are as scalable and potentially operational for field-scale mapping in smallholder systems; and ability to collect accurate plot-level maize yield measurement for hundreds of plots.

Specifically, in Round I, remote sensing based yield estimates were obtained based on four images acquired by the Terra Bella (formerly Skybox) satellites over the 400 km² tasking area on May 15, June 9, June 27, and July 28, 2015. These images were first geometrically corrected to ensure that plot boundaries were properly aligned with the imagery. Clouds and cloud shadows were manually masked out, and each image was radiometrically corrected to surface reflectance by a standard approach of histogram matching to Landsat images from the same locations and time of year. Reflectance values for individual bands were then used to compute two standard vegetation indices, namely (1) normalized difference vegetation index (NDVI) defined by Rouse et al. (1973) as:

$$NDVI = \frac{(Near\ Infrared - Red)}{(Near\ Infrared + Red)}$$

and (2) green chlorophyll vegetation index (GCVI) defined by Gitelson et al. (2003) as:

$$GCVI = \frac{Near\ Infrared}{(Green - 1)}$$

For each image date, the mean NDVI and GCVI values for each plot were calculated based on all pixels that fell within the GPS-based plot boundaries. The crop cutting yields, based on the 20m² area covered by the combination of the 4x4m and the 2x2m subplots, were then used to calibrate an empirical model relating yield to GCVI on the first three imagery dates, namely May 15, June 9, June 27. NDVI was also tested but performed significantly worse.

When using the entire data set of crop-cutting yields on pure stand fields (n = 235), the correlation between yields and GCVI was low, with an adjusted R² (R²_{adj}) of 0.13 from a model using the three dates of GCVI. Based on the possibility that the area covered by the combination of the subplots may not represent the heterogeneity across the entire plot area, the model was estimated using different subsets of plots for which the standard error in the crop-cutting yields (as calculated

based on the standard deviation among the yields obtained from five 2x2 crop cut sub-plots) was below different thresholds. These models demonstrated much higher adjusted R^2 values. For instance, R^2_{adj} was equal to 0.33 when using 20 percentage points as the threshold ($n = 53$) and 0.38 when using 10 percentage points as the threshold ($n = 30$). The latter model was then used to predict yields across the entire plot sample.¹⁷

4.4.2.3. Farmer Estimation

Plot managers were asked to report their estimate of maize harvest at the parcel-plot-level during the post-harvest visit, replicating the design of the relevant Uganda National Panel Survey (UNPS) questionnaire modules.¹⁸ Each plot manager was allowed to report production in non-standard measurement units, and was asked to report on both the condition (e.g. green harvested; dry after additional drying; etc.) and the state (e.g. with cob but without stalk or husk; grain; etc.) of up to three maize harvests that may have occurred on the plot over a period of time. The production measurement units, conditions, and states were borrowed directly from the UNPS, and are provided in Appendix IV. The dry grain-equivalent harvest quantities in kilograms were calculated in each round by using the conversion factor database that has been compiled by the UBOS during the 2007 Uganda Census of Agriculture (UCA) for each non-standard measurement unit-condition-state combination and that has been complemented by the data solicited during the UNPS 2009/10, 2010/11, and 2011/12 waves for the (rare) combinations that were not captured as part of the UCA exercise.¹⁹ Moreover, in Rounds I and II, farmer-reported production information

¹⁷ The MAPS II remote sensing work program was continuing at the time of the publication of the working paper version of this manuscript. Given the limited availability of cloud-free Terra Bella imagery in 2016 (acquired mirroring the timeline of the Round I imagery acquisition), the remote sensing validation in MAPS Round II relies primarily on Sentinel-2 imagery.

¹⁸ It is important to note that the identification of parcels versus plots within parcels was anchored in the precise definitions that have been referenced above and that have been in effect since the UNPS 2009/10 wave. The operationalization of these definitions is such that each enumerator, prior to the administration of the post-planting questionnaire, has a detailed discussion with the holder regarding the organization of his/her farm. This conversation (1) ensures that the enumerator and the farmer are on the same page regarding what parcels versus plots within parcels mean, (2) often culminates in sketches of different parcels and plots within parcels that are being cultivated during that reference season, and (3) establishes how parcels and plots within parcels will be rostered in the questionnaire instrument. The established parcels and plots within parcels are then reviewed at each subsequent visit to the household.

¹⁹ The conversion factors have been made available as part of a study by Oseni et al. (2017), and can be accessed here <https://goo.gl/HgdbBv>. To calculate the dry grain-equivalent harvest quantities, we start with the UCA+UNPS augmented database that contains the kilogram value for each measurement unit-condition-state combination. We merge this file at the unit-condition-state level to each of the reported harvests on the maize plot, and multiply the reported quantity with the conversion factor to calculate an initial kg-equivalence. Further, there are two domains of raw harvest quantities – those that are reported on the cob versus those that are reported as grain. We first work in these domains separately. Within the “cob” domain, we adjust the initial kg-equivalence for all harvest quantities such that they would be reported in terms of the condition “dry, after additional drying” and state “with cob, without husk or stalk” The adjustments are precisely the conversion factor ratios within the cob domain between the preferred condition-state combination and all other observed combinations. Within the “grain” domain, we carry out a similar adjustment procedure such that all harvested quantities are expressed in terms of condition “dry after additional drying” and state “grain.” The final adjustment is for expressing the standardized harvest quantities within the cob domain in terms of condition “dry after additional drying” and state “grain.” To do this, we first compute EA-specific

was solicited for all maize plots cultivated by the household during the respective season, inclusive of the plot on which crop cut sub-plots were laid.

4.4.3. Objective Soil Fertility Measurement (MAPS I Only)

Analysis of soil fertility was done in partnership with the World Agroforestry Center (ICRAF). Plot level soil samples were collected from each plot selected for crop cutting following a protocol carefully designed to maximize the representativeness of the samples while maintaining feasibility of implementation. From each plot, four samples were collected from the top-soil (0-20cm depth) and combined to create one composite top-soil sample. Additionally, a single sub-soil sample (20-50cm depth) was collected from the center of the plot. After being processed at the ICRAF Kampala office, the samples were shipped to ICRAF Nairobi office, where approximately 10 percent were subject to conventional wet chemistry testing and all samples were subject to spectral soil analysis. A portion of this 10 percent sample was used to calibrate the prediction models, while the remainder was used to verify the predictions made onto the spectral data.²⁰ The final results from the soil analysis include key indicators of soil fertility such as pH, texture analysis (% sand, % clay, % silt), cation exchange capacity, and the concentration of multiple elements and macro- and micronutrients, including carbon, nitrogen, and potassium.²¹

4.5. Descriptive Statistics

Before moving to the empirical strategy used to examine the existence of the IR, descriptive statistics are used to illustrate the deviations between yield measures based on farmer-reported maize production vis-a-vis crop-cutting and remote sensing. Though not reported, the average GPS-based plot area was 0.14 hectares in MAPS I (with a maximum value of 1.37 and a standard deviation of 0.16) and 0.18 hectares in MAPS II (with a maximum value of 2.56 and a standard deviation of 0.23).^{22, 23} The plot-level averages for maize production (kilograms) and yield

adjustment factors as the average ratio between shelled and unshelled (on the cob, dry after additional drying, prior to shelling) crop cut sub-plot harvests obtained in MAPS II, and multiply all kg-equivalent production estimates within the “cob” domain in MAPS I and MAPS II with its corresponding EA-specific adjustment factor.

²⁰ For details, see Shepherd and Walsh (2002).

²¹ During the post-planting interview and prior to visiting the plot for GPS-based area measurement and laying crop cut sub-plots, each plot manager was also asked a multitude of questions about the soil attributes, specifically color, texture, type, and overall quality, of the plot selected for crop cutting. We do not provide further information on these variables as we only rely on objectively-measured soil quality index.

²² The round-specific descriptive statistics, including GPS-based plot area, are reported in Table A1. Where available, means are reported for the comparable UNPS 2015/16 sample. Due to high missingness rates of GPS-based area measurement, and area measurement at the parcel level rather than plot level, plot area and yield figures for UNPS are not comparable to those in MAPS I or II, and are, therefore, not reported.

²³ In MAPS I and II, the GPS-based area measure was obtained in each household only for the randomly selected maize plot subject to crop cutting. Farmer-reported area was, however, solicited for the complete set of maize plots that may have been cultivated by the household during the reference agricultural season. Following Kilic et al. (2017a, 2017b), we estimated a simple imputation model of GPS-based plot areas pooled across MAPS I and II as a function of farmer-reported plot area, a dichotomous variable identifying 2016 round observations, and EA fixed effects. The

(kilograms per hectare, using GPS-based plot areas) based on self-reported maize production, crop cutting and remote sensing are presented in Table 2 for each round.²⁴ To quell any concerns of bias in self-reported estimates stemming from exposure to full-plot harvest, the computation of self-reported MAPS II averages excludes plots on which full-plot harvest was conducted. Certain patterns emerge that are consistent with expectations. First, maize production and yield are, on average, higher on pure stand plots than their intercropped counterparts, irrespective of the approach to measurement of maize production. This is part due to still using the entire cultivated plot area, without further adjustments for planting density, while computing the intercropped maize yields. Second, there is a marked decline in productivity from MAPS I to MAPS II. This is in line with the increased incidence of crop damage observed by both plot managers (on the full plot) and enumerators (on the crop cutting sub-plots), primarily attributable to a reported increase in drought.²⁵

Third, in each round, the overall self-reported maize yield is, on average, at least 85 percent higher than its crop cutting and remote-sensing based counterparts. The comparable degree of discrepancy is at least 25 percent in terms of plot-level maize production – although the self-reported maize production is, on average, 84 percent higher than the comparable figure based on full-plot crop cutting. While the means presented in Table 2 illustrate the over-estimation of maize production and yields on the part of our farmers, they do not address whether a systematic bias is at play. The correlates of over-estimation of yields with respect to sub-plot crop cutting based yield measurement are explored in Section 5.

Fourth, the comparison of the averages for production and yield based on alternative measurement methods, including sub-plot crop cutting, full-plot crop cutting and remote sensing, reveals a greater similarity among the estimated yields. On pure stand plots, in MAPS II, the differences in the average plot-level maize yield based on sub-plot crop cutting versus full-plot crop cutting are not statistically significant, lending confidence to sub-plot crop cutting estimates in this domain. Similarly, in MAPS I, the difference in the average plot-level maize yield on pure stand plots based on sub-plot crop cutting versus remote sensing is also not statistically significant, raising hopes regarding the application of high-resolution satellite imagery based remote sensing for the measurement of crop yields in smallholder production systems.

R² for the imputation model was 0.59. Within a multiple imputation framework, we subsequently obtained a single imputation of GPS-based plot area for the maize plots that were not measured, using predictive mean matching. Specifically, we use the linear GPS-based plot area prediction as a distance measure to form a set of 5 nearest neighbors out of the plot sample measured with GPS, and randomly pick one of these neighbors whose observed GPS-based area value replaces the missing value for the incomplete case at hand. The completed data set with the imputed GPS-based plot areas was then collapsed at the household-round-level to induce a better understanding of the scale of the maize farms. The average GPS-based household-level total area cultivated with maize was 0.29 hectares in MAPS I (with a maximum value of 4.49 and a standard deviation of 0.36) and 0.26 hectares in MAPS II (with a maximum value of 2.56 and a standard deviation of 0.26).

²⁴ The sample is restricted to those in which remote sensing, crop-cutting, and self-reported estimates are available, and to those households which had crop-cutting and farmer-estimated production estimates in MAPS I and MAPS II.

²⁵ Please see Table A2 for round-specific breakdowns of the reported reasons for production loss.

On the other hand, our analysis further underscores the difficulty of estimating yields in the intercropped domain, even with objective measurement approaches. In MAPS II, we find that the average full-plot crop cutting yield on intercropped plots is 85 kilograms per hectare lower than its sub-plot crop cutting counterpart, and the difference is significant at the 5 percent level. This is in line with the expectation that the sub-plot crop cutting yield may not entirely reflect the true measure given the across-household variation in the types of crops intercropped with maize, and both across-household and intra-plot variation in maize seeding rate. And, if we can think of sub-plot crop cutting yield estimate in the intercropped domain as an upper bound for the true yield based on MAPS II data, we see that remote sensing overestimates plot-level maize production and yield in MAPS I by a significant margin – likely due to difficulties in attributing vegetation growth to maize production on intercropped plots.²⁶

Overall, the level and direction of discrepancy between farmer estimates of production and crop-cutting observed in MAPS I and II run contrary to some previous works, synthesized by Fermont and Benson (2011). Specifically, while MAPS data reveal that maize yields based on farmer-reported production are significantly over-reported relative to both sub-plot and full-plot crop cutting, Verma et al. (1988) assert that, on average, farmer estimates are in fact more accurate than sub-plot crop-cutting relative to full-plot harvests. There are, however, two concerns associated with this claim.

First, the small sample used by Verma et al. (1988) prohibited analysis at varying plot size levels. Yet, as presented below and in line with the findings of Desiere and Jolliffe (2017), the degree of error between farmer estimates and crop-cutting measures is systematic in nature, with production and yields more significantly over-estimated by the farmer on smaller plots. Ignoring distributional differences between farmer estimates and crop-cutting, therefore, can mask the true relationship between the two measures. Second, the plots analyzed by Verma et al. (1988) were all subject to a full-plot harvest, which would likely contaminate farmer-reported production values. Full-plot harvests also eliminate the need for farmers to aggregate periodic harvests (such as early, green harvest with final, dry harvest), further contributing to the accuracy of farmer-reported production values in the presence of contamination.

Similarly, Fermont and Benson (2011) compile historical maize yield estimates in Uganda from a variety of sources, and report that the self-reported yield estimates are consistently lower than the estimates based on sub-plot crop cutting, in contrast with our findings. Yet, the assertion of Fermont and Benson (2011) is potentially misleading, since their reported estimates based on crop cutting originate from on-farm trials, where farmers may be positively selected and may receive technical assistance for optimal management practices, as opposed to standard household or farm

²⁶ Since the MAPS II remote sensing work program was continuing at the time of the publication of the working paper version of this manuscript, we will refrain from making further concrete statements regarding the accuracy and feasibility of remote sensing for maize yield measurement in smallholder production systems, and will focus on this topic in a separate study.

survey operations in which no agronomic guidance would be provided to the farmers. Hence, crop-cuts from on-farm trials may likely constitute an upper bound for maize yields in field conditions, and would, therefore, be expected to be greater than maize yields based on farmer-reported production.

5. Empirical Strategy

5.1. Inverse Scale-Productivity Relationship

To investigate whether the inverse scale-productivity relationship is sensitive to the way in which the plot-level maize yield is measured, we estimate several variants of three regressions following form:

$$(1) Y_{ih} = \alpha + \beta_1 A_{ih} + \delta P_{ih} + \sigma O_{ih} + \tau H_{ih} + \mu M_{ih} + \varepsilon_{iht}$$

$$(2) Y_{iht} = \alpha + \beta_1 A_{iht} + \delta P_{iht} + \sigma O_{iht} + \tau H_{iht} + \mu M_{iht} + \rho T_{iht} + \theta_{iht} + \varepsilon_{iht}$$

$$(3) Y_{iht} = \alpha + \beta_1 A_{iht} + \delta P_{iht} + \sigma O_{iht} + \tau H_{iht} + \mu M_{iht} + \rho T_{iht} + \gamma_{iht} + \varepsilon_{iht}$$

In each equation, Y is the logarithmic transformation of the plot-level maize yield (kilograms per hectare), based on GPS-based plot area. Equation 1 is a cross-sectional linear regression that is estimated separately in each survey round; Equation 2 is a panel linear regression that is estimated with household fixed effects; Equation 3 is a panel linear regression that is estimated with parcel fixed effects. As shown in Table 3, Equations 1 through 3 are estimated using several plot sample definitions, and alternative maize yield measures based on (1) self-reporting, (2) sub-plot crop cutting, (3) full plot crop cutting, and (4) remote sensing.

The following is an overview of the notation used in Equations 1 through 3. First, i and h denote plot and household, respectively; t denotes survey round in the panel regressions; and α and ε are the constant and the error term, respectively. The common vectors included in all equations include A , P , H , and M , whose choice takes into account the explanatory variables commonly featured in production functions estimated to investigate the inverse scale-productivity relationship.

A is the logarithmic transformation of GPS-based plot area, and β_1 is the main coefficient of interest across all estimations. A negative and statistically significant β_1 would be in support of the inverse scale-productivity relationship at the plot-level.

P is a vector of plot-level characteristics, including (1) a binary variable identifying whether the plot was pure stand with maize; (2) logarithmic transformation of seeding rate under

intercropping²⁷; (3) logarithmic transformation of kilograms of maize seed planted; (4) a binary variable identifying whether any inorganic fertilizer was applied on the plot²⁸; (5) logarithmic transformation of total household member days of season-specific labor input on the plot²⁹; (6) a binary variable identifying whether there was any hired labor input on the plot; (7) logarithmic transformation of total hired days of season-specific labor input on the plot³⁰; (8) percent seasonal (May-June) rainfall deviation from plot location-specific long-term average rainfall³¹; (9) logarithmic transformation of GPS-based distance between plot and dwelling in kilometers; (10) enumerator-assessed percent damage in the crop cut sub-plot³²; and (11) a binary variable identifying whether any cover crops were on the plot prior to planting.

H is a vector of household characteristics, including (1) wealth index, (2) agricultural implement and machinery index, (3) logarithmic transformation of household size, and (4) dependency ratio. M is a vector of plot manager characteristics, including (1) a binary variable identifying whether the respondent is also the plot manager, (2) a binary variable identifying whether the plot manager received agricultural extension services on topics relevant to crop production and marketing in the last 12 months, (3) a binary variable identifying whether the plot manager is female, and (4) logarithmic transformations of plot manager age and years of education.

Specific to Equation 2 and Equation 3, T is our time fixed effect, i.e. binary variable identifying the second round of plot observations in the panel regressions. Further, θ in Equation 2 is the fixed effect that accounts for time-invariant unobserved heterogeneity at the *household-level*, while γ in Equation 3 is the fixed effect that accounts for time-invariant unobserved heterogeneity at the *parcel-level*. The sample for Equation 3 includes plots in both rounds only if the plot selected for crop cutting in MAPS II happened to be on the same *parcel* as the plot selected for crop cutting in MAPS I.³³

Further, the vector O includes the rest of the objectively-measured plot-level covariates, including (1) soil fertility index, (2) genetic heterogeneity of the maize on the plot, based on DNA fingerprinting of seed samples obtained from the combined harvest tied to the 4x4m crop cut sub-

²⁷ This is calculated as the ratio between the quantity of maize seed planted under intercropping and the counterfactual quantity of maize seed that would have been planted had the plot been cultivated pure stand, as reported by the farmer. The pure stand maize plots are assigned a value of 1.

²⁸ The rare occurrence of inorganic fertilizer use prevents us from using the logarithmic transformation of the quantity of inorganic fertilizer applied.

²⁹ This is calculated as the sum of all household member-specific labor inputs reported by the farmer at the plot-level.

³⁰ The plots without any hired labor are assigned a value of 0.

³¹ The plot location-specific dekadal time series rainfall data are sourced from the CHIRPS database. The plot location-specific long-term average for the period of May-June is computed over the period of 1981-2015 for MAPS I, and 1981-2016 for MAPS II.

³² In MAPS I, this is calculated as the average percent damage across both 4x4 and 2x2 sub-plots.

³³ MAPS I parcel rosters were carried forward for updating in MAPS II to identify the MAPS I parcels that were still owned and/or cultivated by the households during the first rainy season of 2016. This information, together with the known nesting of the maize plots with the parcels in each round, was used to determine whether the selected plot in MAPS II happened to be on the same parcel as the selected plot in MAPS I.

plot in MAPS I³⁴, and (3) edge effects, specifically, the share of the crop cut sub-plot that is within 4 meters of the nearest plot edge (and separately, within 1 meter of the nearest plot edge, specifically in Round 2, given the increase in the crop cut sub-plot area to 8x8m). The inclusion of this vector is motivated by the competing hypotheses on the weakening or disappearing inverse scale-productivity relationship as a result of controlling for plot attributes, such as soil quality (Barrett et al., 2010), and edge effects (Bevis and Barrett, 2017). Appendix V provides more details on the construction of the soil fertility index and the edge effects.

5.2. Drivers of Farmer Over-Estimation

Descriptive analysis is first used to highlight the validity of the hypothesized sources of measurement error in self-reported production estimates, primarily heaping of production figures. Consistency of the error across time is additionally explored, with an eye for examining the potential to correct this measurement error econometrically. Following Desiere and Jolliffe (2017), we subsequently move to explore the correlates of farmer over-estimation through regression analysis of the log ratio of self-reported to crop cutting yields. Since it is feasible that the drivers of farmer over-estimation vary along the distribution of over-estimation, recentered influence function (RIF) regressions are estimated. The econometric method, put forth by Firpo et al. (2009), executes an unconditional quantile regression by first estimating an influence function recentered on a given quantile, and subsequently utilizing the estimated RIF value as the dependent variable in a linear regression. The influence function for the dependent variable, Y , in this case the degree of measurement error in self-reported production-based yields proxied by the log ratio of self-reported to crop cutting yields, is as follows:

$$(4) IF(Y; Q_T) = \frac{(T - \mathbf{1}\{Y \leq Q_T\})}{f_Y(Q_T)}$$

where $\mathbf{1}\{Y \leq Q_T\}$ equals 1 if the dependent variable is less than or equal to the quantile Q_T (0 otherwise), Q_T is the population T-quantile of the unconditional distribution of Y , and $f_Y(Q_T)$ is the density of the marginal distribution of Y . The RIF is a function of equation (4), assuming $IF(y; v)$ is the influence function for an observed productivity outcome y :

$$(5) RIF(y; v) = v(F_Y) + IF(y; v)$$

where $v(F_Y)$ is the distributional statistic for the dependent variable. Therefore,

$$(6) RIF(Y; Q_T) = Q_T + IF(Y; Q_T)$$

³⁴ For more information on the MAPS maize variety identification component, please see Ilukor et al. (2017).

Finally, the RIF values are regressed on a series of covariates using linear regressions. The covariates included are those believed to influence productivity itself, as discussed above, as well as controls for whether self-reported production estimates were rounded/heaped on whole numbers, included 100, 200, 300, 400 or 500 kg or 1, 2, 3, 4, 5, or 10 100 kg sacks.

6. Results

6.1. Inverse Scale-Productivity Relationship

Table 4 reports the plot area coefficients from the regressions that are estimated as specified in Table 3. The full set of regression results are reported in the Appendix Tables A4.1-A4.3. As discussed above, a negative and statistically significant coefficient on plot area confirms the IR. The results suggest that when using farmer-reported production estimates to define plot-level maize yields, the IR holds across the entire set of specifications and analysis samples of interest, including in the instances where we control for unobserved time-invariant heterogeneity at the household- and parcel-level. The magnitude of the IR is non-trivial; on average, a 1 percent increase in GPS-based plot area results in a 0.35 to 0.9 percent reduction in maize yields, depending on the specification. In terms of magnitude, the coefficients are similar with respect to other published findings in Uganda (Larson et al. 2014).

On the other hand, there is no evidence of the IR when we define maize yield on the basis of sub-plot crop cutting, full-plot crop cutting, or remote sensing-based crop production estimates. In MAPS II, using full-plot crop cutting maize yield, the GPS-based plot area coefficient is insignificant, statistically indistinguishable from zero, irrespective of cultivation status – suggesting constant returns to scale (CRS). In specifications that exploit the panel nature of the data, the coefficient of interest at times changes direction, but again is insignificant across all specifications and analysis sample definitions, with the exception of a marginally significant, positive coefficient under the parcel-panel column. Similarly, in MAPS I, remote sensing estimates, on average, suggest a marginally significant positive return to land area, but this result is not robust as the coefficient loses significance when the sample is disaggregated by cultivation status, and in the pure stand domain, where we have a higher degree of confidence in remotely-sensed yields, it is near-zero in magnitude and statistically insignificant. Taken together, the findings based on maize yields that are measured with objective survey methods support constant, as opposed to decreasing, returns to scale.

These conclusions hold true (1) in the expanded sample with the households from the two districts, Serere and Sironko, that are excluded from the analysis sample due to lack of remote sensing and variety identification data; (2) if the top and bottom 5 percent of the plots in terms of maize yield are excluded from the analysis sample; and (3) if the analysis sample is limited only to households

that cultivate only one maize plot. These robustness checks, which are available upon request, work to ensure that the results (1) hold across a greater geographic area in the Eastern region; (2) are not being driven by outliers in yield data; and (3) in part rule out the theory that errors in self-reported production are driven by difficulty in reporting production at the plot level, rather than the (maize) farm level.

Further, as documented in the Appendix Table A3, while IR persists at the mean and throughout the distribution of plot-level maize yields anchored in farmer-reported production, CRS is observed throughout the distribution of objective yield measures. This finding is in stark contrast with the new evidence put forth by Savastano and Scadizzo (2017), who, based on farmer-reported panel survey data, provide support for the IR among the sample of farmers above the median land productivity measure.

Contrary also to the findings by Bevis and Barrett (2017), who use the perimeter-to-area ratio as a proxy for edge effect and find plot edges to be more productive than the interior, the results summarized in Table 4, and reported in full in the Appendix Tables A4.1-A4.3, suggest that plot edges are generally not more or less productive than the interior of the plot, and when significant, the direction of the coefficient suggests the plot edges are less productive. This finding is robust to an alternative definition of the edge effect, namely a 1m buffer along the plot edge as opposed to the 4m buffer presented here. The support for CRS, as opposed to the IR, in the MAPS II estimations that use the full-plot crop cut yield as the dependent variable raises further questions on the viability of the edge effect hypothesis in our context.

Finally, to illustrate the difference in plot area coefficients across specifications, Figures 3, 4, and 5 present the plot area coefficients with 95 percent confidence intervals for the overall, pure stand, and intercropped sample, respectively. The confidence intervals around the estimated coefficients originating from regressions that use sub-plot crop cutting-based yield estimates are considerably tighter than the comparable confidence intervals associated with the plot area coefficients estimated from regressions that use self-reported production-based yield estimates. The following section examines the farmer-reported maize production data more closely and seeks to identify the correlates of farmer over-estimation.

6.2. Drivers of Farmer Over-Estimation

Before addressing the correlates of farmer over-estimation and potential sources of systematic error, we first address the feasibility of correcting this error with the use of correction factors. In order for correction factors to be an appropriate course of action to improve the quality of self-reported production data, the error must be consistent across time and within household. Figure 6, however, illustrates that this is not the case.

The pairwise correlation of the ratio of self-reported to sub-plot crop cut production estimates from MAPS I to MAPS II is 0.002, not statistically distinguishable from zero. Confidence in the consistency of self-reported measurement error in crop production is further degraded by the lack of predictive power from year to year. Table 5 presents the results of simple OLS regression of MAPS II yields on MAPS I yields, using both sub-plot crop cut production (columns 1, 2, and 3) and self-reported production (columns 4, 5, and 6).

When relying on the objective measure of production in the numerator, MAPS I yield is a significant predictor of MAPS II yields, particularly in the sample in which both plots were from the same parcel. This reflects farmer ability, the quality of the land, agricultural practices employed, etc. Conversely, self-reported production-based yields in MAPS I are not a statistically significant predictor of yields in MAPS II. The correlation of self-reported yields between MAPS I and II on the parcel panel sample is only 0.012, while the correlation between sub-plot crop cutting yields is 0.226. The inconsistency of measurement error in self-reported production estimates observed in Figure 6 and Table 5 severely limits the potential for utilizing a “corrected” self-reported production data in productivity analysis.

If the over-estimation of maize yields based on farmer-reported maize production were uniform across the distribution of plot sizes, it would have no influence on the observed relationship between land productivity and plot area. Rather, the differences in the plot area coefficients suggest that there is a systematic bias in farmer-reported maize production. To address whether the deviation between crop cutting and farmer estimates varies across the distribution of plot areas, Figure 7 presents the mean deviation in yields (as measured by self-reported yield minus crop-cutting yield) by GPS-based plot area quintile for MAPS I and MAPS II. A clear downward trend exists, in which farmers more greatly over-estimate yields on smaller plots, with the degree of over-estimation decreasing with plot area. This systematic bias would have direct consequences on the IR.

This trend is consistent with evidence on the systematic nature of measurement error in farmer self-reported estimates of plot area. Carletto et al. (2013, 2015, Forthcoming) find that farmers overestimate plot areas relative to GPS-based area measurements more significantly on smaller plots, while only slightly overestimating (or underestimating) the area of larger plots. The collection of both GPS-based area measurement and farmer self-reported area in MAPS I and II allows for analysis of the relationship between the measurement error in self-reported crop production and plot area. As expected, the degree of measurement error observed in these two variables is significantly and positively correlated, with a correlation coefficient of 0.437 and 0.535 in MAPS I and II, respectively.³⁵

³⁵ The pairwise correlation coefficient of the ratio of self-reported to sub-plot crop cut based production to the ratio of self-reported to GPS-based plot area measurement is reported. In MAPS II, plots which were subject to full-plot crop cutting are excluded.

To explore another potential source of farmer over-estimation yield, namely the persistence of rounding of farmer-reported production estimates, Figure 8 presents a histogram of reported quantities for the primary production units, kilograms and 100 kilogram sacks for MAPS I and MAPS II, as well as the maize-producing UNPS 2015/16 sample residing the Eastern region and reporting for the first season of 2015 (overlapping with the MAPS I reference period). There is clear evidence of heaping at common intervals, such as 50 kg, 100 kg, 200 kg, and 300 kg, and 1 sack, 2 sacks, and 10 sacks, which may explain, at least in part, the trend in greater over-estimation on smaller plots. Rounding of self-reported production quantity on plots with lower production may have a more severe impact in terms of percent of total production, therefore resulting in greater bias.

The results of the RIF estimations, as described in section 4, are presented in Table 6. Plot area is indeed negatively correlated with farmer over-estimation in MAPS I across the entire distribution of over-estimation, while area is only significantly different from zero (and negative) in the upper tail for MAPS II and the household panel. Supporting the theory that subjective estimates of production are complicated by multiple harvest periods and crop conditions, the number of conditions in which the farmer reports overall harvest is positively correlated with over-estimation in MAPS I and the household panel (insignificant in MAPS II except for the least productive decile). Also exacerbating yield over-estimation are the farmer-reported seed, household labor and hired labor input quantities, which may, therefore, suffer from the same subjective sources of measurement error. This theory is supported by the correlation of measurement error observed between self-reported plot area and crop production discussed above, and signals potential concern for production function estimations more broadly.

Finally, the enumerator-assessed percent damage in the crop cut sub-plot, which could be taken as a proxy to the damage on the plot as a whole, is significantly and positively correlated with farmer over-estimation, suggesting farmers may not be taking into account crop damage when estimating plot-level production. While some may theorize that farmer over-estimation originates in the inability of farmers to distinguish between the harvest on multiple plots (that is, potentially reporting the production of multiple plots on a single plot), the results in Table 6 suggest otherwise. Whether the household cultivates more than one maize plot or not has no bearing on the degree of farmer over-estimation.

7. Conclusions

Based on a two-round household panel survey conducted in Eastern Uganda to test the relative accuracy of subjective approaches to data collection vis-à-vis objective survey methods for maize yield measurement, soil fertility assessment, and maize variety identification, we provide unambiguous support for the sensitivity of the plot-level inverse scale-productivity relationship to

the choice of the method by which maize production and yield (anchored in GPS-based plot area measurement) are computed.

While farmer-reported production-based plot-level maize yield regressions consistently lend support to the inverse scale-productivity relationship, in magnitudes that are similar to previously published findings on Uganda and other African settings, the comparable regressions estimated with maize yields based on sub-plot crop cutting, full-plot crop cutting, and remote sensing point towards constant returns to scale (CRS). In view of the competing hypotheses for the IR, the regressions control for objective measures of soil fertility, maize genetic heterogeneity, and edge effects at the plot-level; a rich set of plot, household and plot manager attributes; as well as household and parcel fixed effects in select specifications. The existence of the IR anchored in the use of farmer-reported plot-level maize production is also shown throughout the yield distribution, while CRS is documented throughout the distribution of maize yields based on crop cutting and remote sensing variants.

Our core finding is driven by persistent over-estimation of farmer-reported maize production and yield vis-à-vis their crop cutting-based counterparts, particularly in the lower half of the plot area distribution. The analysis (1) points to the rounding of maize production as a key factor in farmer over-estimation, (2) suggests that farmers do not consider the degree of crop damage when reporting production, and (3) provides evidence for multiple harvests (and the accompanying heterogeneity in harvested crop conditions) contributing to over-estimation. While some may argue that farmer over-estimation is driven by the inability of farmers to distinguish production at the plot level, rather than farm level, our results also serve to refute this claim.

Though the findings contribute to a larger (and renewed) body of literature questioning the inverse scale-productivity relationship based on omitted explanatory variables or alternative formulations of the agricultural productivity measures, the analysis, together with Desiere and Jolliffe (2017), is among the first documenting how errors in self-reported survey data on production mediate the existence of the IR. The consistency in the findings across our study and that of Desiere and Jolliffe (2017), which has a focus on Ethiopia, is noteworthy.

Taking into account the similarities in heaping in self-reported maize production information across our study and the Eastern region sub-sample of the UNPS for the same agricultural season, our findings emphasize the need for sustained focus on the improvement of crop production and yield measurement in the context of household and farm surveys that solicit farmer-reported production information and that capture farming at similar scales as well as pervasive use of non-standard measurement units, and heterogeneity in harvested crop conditions that could be reported across and within households.

Although we use the official maize unit-condition-state specific conversion factors that are further augmented with MAPS-based measurements that provide additional nuance to expression of unshelled maize in shelled equivalent terms, the quality of conversion factors that are used for computation of farmer-reported maize production in kg-equivalent terms may mediate the accuracy of the self-reported maize production. However, there are open empirical questions regarding the extent that conversion factors should be specific in spatial and temporal terms, and whether an improved set of conversion factors, including through the introduction of non-standard measurement units that recognize the variation in sizes for specific units, would be enough to overcome other challenges that may plague self-reported production information, as reviewed in Section 2.

Finally, given the absence of the IR based on the objective productivity and plot area measures in our sample, the results provide further support for promoting a policy environment that reduces the scope for further subdivision of land and that prioritizes investments in land titling and land market development. It is important to note, however, that our findings do not suggest a broad development policy focus shift away from the needs of smallholder farming households since there are social and economic reasons for why one inherently cares about the improvement of living standards across a large segment of the population relying on farming. And to the extent that cultivating small plot(s) is positively correlated with poverty and that the smallest plots are associated with the most upward bias, the actual yields attained by the poorest may be less than previously estimated.

References

- Ali, D. A., and Deininger, K. (2015). Is There a Farm Size-Productivity Relationship in African Agriculture? Evidence from Rwanda.” *Land Economics*, 91, pp. 317–343.
- Andrews, S.S., Mitchel, J.P., Mancinelli, R., Karlen, D.L., Hartz, T.K., et al. (2002). “On-Farm Assessment of Soil Quality in California’s Central Valley.” *Agronomy Journal*, 94, pp. 12–23.
- Assuncao, J., and Ghatak, M. (2003). “Can unobserved heterogeneity in farmer ability explain the inverse relationship between farm size and productivity?” *Economics Letters*, 80, pp. 189–194.
- Assuncao, J., and Braido, L.H. (2007). “Testing household-specific explanations for the inverse productivity relationship.” *American Journal of Agricultural Economics*, 89, pp. 980–990.
- Atzberger, C. (2013). “Advances in remote sensing of agriculture: context description, existing operational monitoring systems and major information needs.” *Remote Sensing*, 5, pp. 949–981.
- GoU (2013). “Uganda vision 2040.” Retrieved on May 11, 2017 from <https://goo.gl/DEZwNn>.
- Barrett, C. (1996). “On price risk and the inverse farm size–productivity relationship.” *Journal of Development Economics*, 51, pp. 193–215.
- Barrett, C.B., Bellemare, M.F., and Hou, J.Y. (2010). “Reconsidering conventional explanations of the inverse productivity-size relationship.” *World Development*, 38, pp. 88–97.
- Bevis, L., and Barrett, C. B. (2017). “Close to the edge: do behavioral explanations account for the inverse productivity relationship?” Retrieved on August 2, 2017 from <https://goo.gl/j9dyJM>.
- Becker-Reshef, I., Vermote, E., Lindeman, M., and Justice, C. (2010). “A generalized regression-based model for forecasting winter wheat yields in Kansas and Ukraine using MODIS data.” *Remote Sensing of Environment*, 114, pp. 1312–1323.
- Benjamin, D. (1995). “Can unobserved land quality explain the inverse productivity relationship?” *Journal of Development Economics*, 46, pp. 51–84.
- Bhalla, S., and Roy, P. (1988). “Mis-specification in farm productivity analysis: the role of land quality.” *Oxford Economic Papers*, 40, pp. 55–73.
- Carletto, G., Gourlay, S., Murray, S. and Zezza, A. (Forthcoming). “Cheaper, faster and more than good enough: is GPS the new gold standard in land area measurement?” *Survey Research Methods*.
- Carletto, G., Gourlay, S., and Winters, P. (2015). “From guesstimates to GPStimates: land area measurement and implications for agricultural analysis.” *Journal of African Economies*, 24, pp. 593–628.
- Carletto, G., Savastano, S., and Zezza, A. (2013). “Fact or artifact: the impact of measurement errors on the farm size–productivity relationship.” *Journal of Development Economics*, 103(C), pp. 254–261.

- Davis, B., Di Giuseppe, S., and Zezza, A. (2017). "Are African households (not) leaving agriculture? Patterns of households' income sources in rural Sub-Saharan Africa." *Food Policy*, 67, pp. 153-174.
- Desiere, S. and D. Jolliffe. (2017). "Land productivity and plot size: is measurement error driving the inverse relationship?" World Bank Policy Research Paper No. 8134.
- Dorosh, P., and Thurlow, J. (2016). "Beyond agriculture *versus* non-agriculture: decomposing sectoral growth-poverty linkages in five African countries." *World Development*.
- Eastwood, R., Lipton, M., and Newell, A. (2010). "Farm size," In R. Evenson and P. Pingali (Eds.), *Handbook of Agricultural Economics*, Volume 4, Amsterdam: Elsevier.
- Eswaran, M., and Kotwal, A. (1986). "Access to capital and agrarian production organization." *Economic Journal*, 96, pp. 482–498.
- Fermont, A., and Benson, T. (2011). "Estimating yield of food crops grown by smallholder farmers: a review in the Uganda context." International Food Policy Research Institute Discussion Paper No. 1097.
- Franch, B., Vermote, E. F., Becker-Reshef, I., Claverie, M., Huang, J., Zhang, J., Justice, C. and Sobrino, J. A. (2015). "Improving the timeliness of winter wheat production forecast in the United States of America, Ukraine and China using MODIS data and NCAR Growing Degree Day information." *Remote Sensing of Environment*, 161, pp. 131–148.
- Firpo, S., Fortin, N. M., and Lemieux, T. (2009). "Unconditional quantile regressions." *Econometrica*, 77, 953-973.
- Gallego, J., Carfagna, E., and Baruth, B. (2010). "Accuracy, objectivity and efficiency of remote sensing for agricultural statistics agricultural survey methods." In R. Benedetti, M. Bee, G. Espa and F. Piersimoni (Eds.), *Agricultural Survey Methods*, Chichester, UK: John Wiley & Sons, Ltd.
- Gitelson, A. A., Viña, A., Arkebauer, T. J., Rundquist, D. C., Keydan, G., and Leavitt, B. (2003). "Remote estimation of leaf area index and green leaf biomass in maize canopies." *Geophysical Research Letters*, 30, pp. 1248.
- Ilukor, J., Kilic, T., Stevenson, J., Gourlay, S., Kosmowski, F., Kilian, A., Sserumaga, J., and Asea, G. (2017). "Blowing in the Wind: The Quest for Accurate Crop Variety Identification in Field Research, with an Application to Maize in Uganda." *Mimeo*.
- Kilic, T., Palacios-Lopez, and Goldstein, M. (2015a). "Caught in a productivity trap: a distributional perspective on gender differences in Malawian agriculture." *World Development*, 70, pp. 416-463.
- Kilic, T., Winters, P., and Carletto, C. (2015b). "Gender and agriculture in Sub-Saharan Africa: introduction to the special issue." *Agricultural Economics*, 46, pp. 281-284.
- Kilic, T., Yacobou Djima, I. and Carletto, C. (2017a). "Mission impossible? Exploring the promise of multiple imputation for predicting missing GPS-based land area measures in household surveys." World Bank Policy Research Working Paper No. 8138.

- Kilic, T., Zezza, A., Carletto, C., and Savastano, S. (2017b). “Missing(ness) in action: selectivity bias in GPS-based land area measurements.” *World Development*, 92, pp. 143-157.
- Lamb, R. L. (2003). “Inverse productivity: land quality, labor markets, and measurement error.” *Journal of Development Economics*, 71, pp. 71–95.
- Larson, D. F., Otsuka, K., Matsumoto, T., and Kilic, T. (2014). “Should African rural development strategies depend on smallholder farms? An exploration of the inverse-productivity hypothesis.” *Agricultural Economics*, 45, pp. 355–367.
- Lobell, D. B. (2013). “The use of satellite data for crop yield gap analysis.” *Field Crops Research*, 143, pp. 56–64.
- Lobell, D. B., Asner, G. P., Ortiz-Monasterio, J. I., and Benning, T. L. (2003). “Remote sensing of regional crop production in the Yaqui Valley, Mexico: estimates and uncertainties.” *Agriculture, Ecosystems & Environment*, 94, pp. 205-220.
- Lobell, D. B., Thau, D., Seifert, C., Engle, E., and Little, B. (2015). “A scalable satellite-based crop yield mapper.” *Remote Sensing of Environment*, 164, pp. 324–333.
- Lowder, S. K., Skoet, J., and Raney, T. (2016). “The number, size, and distribution of farms, smallholder farms, and family farms worldwide.” *World Development*, 87, pp. 16-29.
- MAAIF (2013). “National agriculture policy.” Uganda, Kampala: MAAIF.
- Mukherjee A., and Lal, R. (2014). “Comparison of soil quality index using three methods.” *PLoS ONE*, 9.
- Nkonya, E., Pender, J., Jagger, P., Sserunkuuma, D., Kaizzi, C., and Ssali, H. (2004). “Strategies for sustainable land management and poverty reduction in Uganda.” International Food Policy Research Institute Research Report No. 133.
- Oseni, G., Durazo, J., and McGee, K. (2017). “The use of non-standard measurement units for the collection of food quantity: a guidebook for improving the measurement of food consumption and agricultural production in living standards surveys.” LSMS Guidebook, Washington, DC: World Bank. Retrieved on August 7, 2017 from <https://goo.gl/DMK97z>.
- Rouse, J. W., Haas, R. H., Schell, J. A., and Deering, D. W. (1973). “Monitoring vegetation systems in the great plains with ERTS.” In *Third ERTS Symposium* (Vol. I, pp. 309–317). NASA SP-351.
- Savastano, S., and Scandizzo, P. L. (2017). “Farm size and productivity: a ‘direct-inverse-direct’ relationship.” World Bank Policy Research Working Paper No. 8127.
- Shanahan, J. F., Schepers, J. S., Francis, D. D., Varvel, G. E., Wilhelm, W. W., Tringe, J. M., Schlemmer, M. R., and Major, D. J. (2001). “Use of remote-sensing imagery to estimate corn grain yield.” *Agronomy Journal*, 93, pp. 583–589.
- Sibley, A. M., Grassini, P., Thomas, N. E., Cassman, K. G., and Lobell, D. B. (2014). “Testing remote sensing approaches for assessing yield variability among maize fields.” *Agronomy Journal*, 106, pp. 24-32.
- Uganda Bureau of Statistics (UBOS) (2010). “Uganda census of agriculture 2008/2009 - crop area and production report.” Retrieved on May 11, 2017 from <https://goo.gl/hxn2MS>.

- Verma, V., Marchant, T., and Scott, C. (1988). *Evaluation of crop-cut methods and farmer reports for estimating crop production: results of a methodological study in five African countries*. London: Longacre Agricultural Development Centre Ltd.
- World Bank (2016). "The Uganda poverty assessment report. Farms, cities and good fortune: assessing poverty reduction in Uganda from 2006 to 2013." Retrieved on May 11, 2017 from <https://goo.gl/Pvp9ew>.

Tables

Table 1. Potential Sources of Error in Farmer-Reported Estimates of Production

<i>Type</i>	<i>Mechanisms</i>
Measurement Complications	Conversion of non-standard production units to standard units or monetary values Variation in crop condition and state at harvest
Unintentional Bias	Recall bias Rounding of production quantity Partial early/green harvest Extended harvest/permanent crops
Intentional Bias	Perceived benefits of under-reporting (such as eligibility for program or service) Desire to appear successful, leading to over-reporting

Table 2. Mean Plot-Level Maize Yield (Kg/Ha) and Production (Kg) by Round, Survey Method & Cultivation Pattern & Results from Tests of Mean & Distributional Differences

	MAPS I			MAPS II		
	All	Pure Stand	Intercropped	All	Pure Stand	Intercropped
Yields (kg/ha)						
Self-Reported (SR)	2,634	3,035	2,296	1,838	1,822	1,845
Sub-Plot Crop Cutting (CC)	1,064	1,220	932	731	835	693
Remote Sensing (RS)	1,425	1,191	1,622	N/A	N/A	N/A
Full Plot Crop Cutting (FP)	N/A	N/A	N/A	665	854	608
Difference in Means:						
<i>SR vs CC</i>	***	***	***	***	***	**
<i>SR vs RS</i>	***	***	-	N/A	N/A	N/A
<i>CC vs FP</i>	N/A	N/A	N/A	**	-	**
<i>CC vs RS</i>	***	-	***	N/A	N/A	N/A
Difference in Distributions:						
<i>SR vs CC</i>	***	***	***	***	***	***
<i>SR vs RS</i>	***	***	***	N/A	N/A	N/A
<i>CC vs FP</i>	N/A	N/A	N/A	***	***	***
<i>CC vs RS</i>	***	***	***	N/A	N/A	N/A
Quantity Harvested per Plot (kg)						
Self-Reported (SR)	242	319	177	212	293	178
Sub-Plot Crop Cutting (CC)	173	234	122	145	205	123
Remote Sensing (RS)	191	186	195	N/A	N/A	N/A
Full Plot Crop Cutting (FP)	N/A	N/A	N/A	115	211	85
Difference in Means:						
<i>SR vs CC</i>	***	***	***	***	***	**
<i>SR vs RS</i>	***	***	-	N/A	N/A	N/A
<i>CC vs FP</i>	N/A	N/A	N/A	***	-	***
<i>CC vs RS</i>	-	**	***	N/A	N/A	N/A
Difference in Distributions:						
<i>SR vs CC</i>	***	***	***	***	***	***
<i>SR vs RS</i>	***	***	***	N/A	N/A	N/A
<i>CC vs FP</i>	N/A	N/A	N/A	***	***	***
<i>CC vs RS</i>	***	***	***	N/A	N/A	N/A

Notes: The sample is limited to households with crop cutting in both MAPS I and MAPS II. MAPS II self-reported figures further exclude observations subject to full-plot harvest. ***/**/* denote statistical significance at the 1/5/10 percent level, respectively. The distributional differences are assessed based on the Kolmogorov–Smirnov tests of the equality of distributions.

Table 3. Overview of Cross-Sectional & Panel Regressions Estimated with Alternative Sample & Dependent Variable Definitions

<i>Cultivation Domain</i>	<i>Source of Dependent Variable Maize Yield (Kg/Ha)</i>	Observation Numbers by Regression Type			
		<i>MAPS I Cross-Sectional</i>	<i>MAPS II Cross-Sectional</i>	<i>Household Panel</i>	<i>Parcel Panel</i>
All	Self-Reporting‡	440	240	480	316
	Sub-Plot Crop Cutting‡	N/A	240	480	316
	Sub-Plot Crop Cutting (All Plots)	440	440	880	594
	Remote Sensing	440	In Progress	In Progress	In Progress
	Full Plot Crop Cutting	N/A	200	N/A	N/A
Pure Stand	Self-Reporting‡	201	70	118	64
	Sub-Plot Crop Cutting‡	N/A	70	118	64
	Sub-Plot Crop Cutting (All Plots)	201	117	186	114
	Remote Sensing	201	In Progress	In Progress	In Progress
	Full Plot Crop Cutting	N/A	47	N/A	N/A
Intercropped	Self-Reporting‡	239	170	240	170
	Sub-Plot Crop Cutting‡	N/A	170	240	170
	Sub-Plot Crop Cutting (All Plots)	239	323	430	312
	Remote Sensing	239	In Progress	In Progress	In Progress
	Full Plot Crop Cutting	N/A	153	N/A	N/A

Notes: ‡ MAPS II cross-sectional regressions exclude plots subject to full-plot crop cutting.

Table 4. Overview of Log Plot Area (GPS, Ha) Regression Coefficients

<i>Cultivation Domain</i>	<i>Source of Dependent Variable Log Maize Yield (Kg/Ha)</i>	<i>MAPS I Cross-Sectional</i>	<i>MAPS II Cross-Sectional</i>	<i>Household Panel</i>	<i>Parcel Panel</i>
All	Self-Reporting‡	-0.708***	-0.731***	-0.730***	-0.790***
	Sub-Plot Crop Cutting‡	N/A	0.161*	0.063	0.116
	Sub-Plot Crop Cutting (All Plots)	0.066	0.138**	0.093	0.155*
	Remote Sensing	0.273*	In Progress	In Progress	In Progress
	Full Plot Crop Cutting	N/A	-0.093	N/A	N/A
Pure Stand	Self-Reporting‡	-0.561***	-0.805***	-0.346*	-0.890***
	Sub-Plot Crop Cutting‡	N/A	-0.169	0.084	-0.124
	Sub-Plot Crop Cutting (All Plots)	0.159**	-0.071	-0.092	-0.126
	Remote Sensing	0.047	In Progress	In Progress	In Progress
	Full Plot Crop Cutting	N/A	0.018	N/A	N/A
Intercropped	Self-Reporting‡	-0.829***	-0.702***	-0.705***	-0.872***
	Sub-Plot Crop Cutting‡	N/A	0.260**	0.146	0.208
	Sub-Plot Crop Cutting (All Plots)	-0.087	0.201**	0.117	0.148
	Remote Sensing	0.470	In Progress	In Progress	In Progress
	Full Plot Crop Cutting	N/A	-0.101	N/A	N/A

Notes: ‡ MAPS II cross-sectional regressions exclude plots subject to full-plot crop cutting. MAPS I regressions controls for plot-level (1) soil quality index, (2) maize genetic heterogeneity, and (3) edge effects (specifications 2, 4, and 6). MAPS II regressions based on self-reported maize yield exclude observations subject to full-plot harvest. ***/**/* denote statistical significance at the 1/5/10 percent level, respectively.

Table 5. Predictive Power of MAPS I Yields

<i>Dependent Variable:</i>	CC Yield MAPS II			SR Yield MAPS II‡		
	<i>All</i>	<i>Parcel Panel</i>	<i>Not Parcel Panel</i>	<i>All</i>	<i>Parcel Panel</i>	<i>Not Parcel Panel</i>
<i>Sample:</i>						
CC Yield MAPS I	0.121**	0.143***	0.088	-0.002	0.007	-0.092
SR Yield MAPS I						
Constant	582.775***	545.015***	647.650***	1849.667***	1566.281***	2504.644***
N	240	158	82	240	158	82
R²	0.034	0.051	0.016	0.000	0.000	0.001

Notes: Robust standard errors; ‡ MAPS II cross-sectional regressions exclude plots subject to full-plot crop cutting.

Correlation coefficients:

Parcel Panel Only; excludes households with full plot crop-cut in MAPS II

	CC I	CC II	SR I	SR II
CC MAPS I	1			
CC MAPS II	0.2263	1		
SR MAPS I	-0.0085	-0.0330	1	
SR MAPS II	0.0029	0.0251	0.0116	1

Table 6. Regression Results on Farmer Yield Over-Estimation
Dependent Variable: Log of Plot-Level Ratio of Self-Reported to Sub-Plot Crop Cutting Yield

<i>Decile</i>	MAPS I								
	<i>10</i>	<i>20</i>	<i>30</i>	<i>40</i>	<i>50</i>	<i>60</i>	<i>70</i>	<i>80</i>	<i>90</i>
<u><i>Plot Characteristics</i></u>									
Log Plot Area (GPS, ha)	-0.359*	-0.221**	-0.341***	-0.234**	-0.317***	-0.499***	-0.365***	-0.296**	-0.856***
Production Heaped†	0.658***	0.351***	0.273**	0.267**	0.370***	0.311*	0.258	0.507***	0.684**
Log Plot Distance from Dwelling (GPS, km)	0.120	0.023	0.037	-0.011	0.032	0.111	0.173*	0.016	0.007
Pure stand†	0.172	0.073	0.050	0.131	0.169	0.423**	0.070	-0.014	0.073
% Enumerator-Assessed Damage in Crop Cut Sub-Plot	0.002	0.002	0.004*	0.006**	0.009***	0.012***	0.017***	0.015***	0.023***
Log Intercropping Seeding Rate (=100 for Pure stand Plots)	-0.568*	-0.110	-0.044	-0.209	-0.012	-0.134	0.099	-0.147	0.055
Cover Crops Present Prior to Planting†	0.074	-0.099	0.060	-0.041	-0.052	0.088	-0.125	-0.100	-0.279
Log Maize Seed Planting Rate (Kg/Ha)	0.189	0.122*	0.060	0.124*	0.220**	0.182	0.209*	0.240**	0.172
Organic Fertilizer Application†	0.653	-0.059	0.255	0.518	-0.406	-0.232	0.241	-0.046	-0.945*
Inorganic Fertilizer Application†	-0.218	-0.032	-0.098	-0.015	-0.056	-0.056	-0.077	0.042	-0.154
Log Household Labor Days	-0.075	-0.026	-0.040	-0.019	-0.068	0.024	0.120	0.085	0.067
Log Hired Labor Days	0.066	0.052	0.105	0.153**	0.277***	0.283***	0.325***	0.287***	0.260
No Hired Labor†	-0.065	-0.012	0.129	0.328	0.832**	0.784*	0.970**	0.673	0.258
% Deviation of Seasonal (May-June) Rainfall from Long-Term Mean	0.016	0.000	0.007	0.017	0.016	0.028	0.043**	0.033*	0.007
Harvest Reported in More Than One Condition†	0.794***	0.362***	0.313**	0.299**	0.430***	0.425**	0.339*	0.231	-0.013
<u><i>Household Characteristics</i></u>									
Wealth Index	0.098	-0.029	0.053	0.081	0.036	0.124	0.029	-0.087	-0.205
Agricultural Asset Index	0.091	0.094*	0.058	0.017	-0.100	-0.165	-0.079	0.020	0.050
Dependency Ratio	0.062	0.010	-0.042	-0.082	-0.083	-0.140*	-0.115	-0.151**	-0.124
Log Household Size	0.108	0.120	0.082	0.086	0.143	0.044	-0.258	-0.177	-0.110
HH Cultivates More Than One Maize Plot	-0.133	-0.227**	-0.190	0.026	0.042	-0.001	-0.113	-0.143	-0.383
<u><i>Manager Characteristics</i></u>									
Manager = Respondent†	-0.283	-0.107	0.001	0.011	0.029	0.119	-0.056	0.151	-0.285
Received Crop-Production Related Extension Services†	0.086	0.174*	0.046	0.060	0.138	0.018	-0.028	-0.016	0.157
Female†	-0.280	-0.004	-0.081	-0.040	-0.182	-0.223	-0.257	-0.163	-0.249
Log Age (Years)	0.172	-0.011	0.127	-0.026	-0.036	-0.081	0.025	0.221	0.403
Log Years of Education	0.061	0.024	-0.024	-0.004	-0.043	-0.109	-0.082	-0.060	-0.098
Round II Indicator (=1) †									
Constant	-2.638	-1.814*	-2.684**	-1.983*	-3.501**	-3.590*	-4.744***	-3.692**	-4.308
Enumerator Dummies?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Random Effects?	-	-	-	-	-	-	-	-	-
N	440	440	440	440	440	440	440	440	440
R²	0.156	0.191	0.165	0.201	0.253	0.289	0.273	0.261	0.221

Continued next page

Table 6. Regression Results on Farmer Yield Over-Estimation (continued)
Dependent Variable: Log of Plot-Level Ratio of Self-Reported to Sub-Plot Crop Cutting Yield

<i>Decile</i>	MAPS II‡								
	<i>10</i>	<i>20</i>	<i>30</i>	<i>40</i>	<i>50</i>	<i>60</i>	<i>70</i>	<i>80</i>	<i>90</i>
<u><i>Plot Characteristics</i></u>									
Log Plot Area (GPS, ha)	-0.036	0.030	0.011	0.012	0.048	-0.215	-0.235**	-0.542**	-0.495
Production Heaped†	0.245	0.386	0.329*	0.323**	0.504***	0.977***	0.315	0.633*	0.705
Log Plot Distance from Dwelling (GPS, km)	-0.107	-0.123	0.010	-0.005	-0.041	-0.105	-0.016	0.062	0.118
Pure stand†	0.176	0.193	0.293*	0.088	0.083	0.007	0.061	0.011	0.580
% Enumerator-Assessed Damage in Crop Cut Sub-Plot	0.016**	0.021***	0.013***	0.012***	0.012***	0.022***	0.014***	0.024***	0.024**
Log Intercropping Seeding Rate (=100 for Pure stand Plots)	0.365*	0.042	0.032	-0.006	-0.077	-0.141	-0.075	-0.287	-0.934*
Cover Crops Present Prior to Planting†	0.357	-0.049	-0.241	-0.112	-0.228	-0.219	0.152	-0.029	-0.909*
Log Maize Seed Planting Rate (Kg/Ha)	0.306	0.438***	0.325***	0.251***	0.305***	0.487***	0.424***	0.737***	0.701**
Organic Fertilizer Application†	-0.198	-0.445	0.042	-0.713*	-0.489	-0.279	-0.027	-0.635	-0.817
Inorganic Fertilizer Application†	0.171	0.296	0.025	0.200	0.248	0.184	-0.249	-0.581	-0.165
Log Household Labor Days	-0.026	0.118	0.143*	0.103	0.140**	0.084	0.025	0.054	0.722***
Log Hired Labor Days	0.235	0.127	0.111	0.174**	0.116	0.102	0.109	0.407*	1.157***
No Hired Labor†	0.746	0.371	0.330	0.529	0.309	0.303	0.177	0.827	2.839**
% Deviation of Seasonal (May-June) Rainfall from Long-Term Mean	-0.028	-0.066	0.007	0.019	0.035	0.082	0.047	0.091	-0.174
Harvest Reported in More Than One Condition†	0.717**	0.332	0.115	0.119	0.051	-0.025	-0.036	-0.192	-0.429
<u><i>Household Characteristics</i></u>									
Wealth Index	0.045	0.202	0.027	0.043	0.015	0.094	0.020	0.094	0.343
Agricultural Asset Index	0.135	0.089	0.055	0.051	0.052	-0.060	0.007	-0.139	-0.028
Dependency Ratio	0.019	0.175	0.064	0.061	0.085	0.018	0.036	-0.037	-0.218
Log Household Size	0.543**	-0.134	-0.044	-0.001	0.076	0.150	0.156	0.304	0.004
HH Cultivates More Than One Maize Plot	0.248	0.441**	0.051	0.111	0.028	-0.028	-0.291*	-0.171	-0.266
<u><i>Manager Characteristics</i></u>									
Manager = Respondent†	0.332	0.066	-0.070	0.026	0.025	-0.045	-0.115	-0.664*	-0.183
Received Crop-Production Related Extension Services†	-0.379	-0.477**	-0.154	0.048	0.136	0.381	0.318*	0.420	0.029
Female†	-0.218	-0.161	-0.083	0.027	-0.073	0.051	0.167	0.807**	0.894*
Log Age (Years)	-0.397	-0.096	0.087	0.049	0.062	-0.023	0.015	-0.343	0.006
Log Years of Education	0.114	0.216	0.146	0.049	0.003	-0.080	-0.077	-0.178	-0.214
Round II Indicator (=1) †									
Constant	-4.898**	-3.770*	-3.598***	-2.933***	-2.639***	-3.240*	-2.019	-1.884	-3.182
Enumerator Dummies?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Random Effects?	-	-	-	-	-	-	-	-	-
N	237	237	237	237	237	237	237	237	237
R²	0.204	0.239	0.257	0.239	0.292	0.276	0.286	0.333	0.358

Continued next page

Table 6. Regression Results on Farmer Yield Over-Estimation (continued)
Dependent Variable: Log of Plot-Level Ratio of Self-Reported to Sub-Plot Crop Cutting Yield

Decile	Household Panel‡								
	10	20	30	40	50	60	70	80	90
<u>Plot Characteristics</u>									
Log Plot Area (GPS, ha)	-0.153	-0.110	-0.166*	-0.061	-0.094	-0.326**	-0.394***	-0.465***	-0.427**
Production Heaped†	0.479***	0.373***	0.372***	0.331***	0.505***	0.551***	0.221	0.433**	0.454*
Log Plot Distance from Dwelling (GPS, km)	0.032	0.006	0.008	0.014	0.043	0.041	0.055	0.007	-0.054
Pure stand†	0.137	0.215*	0.211	0.155	0.097	0.203	0.201	0.287	0.444*
% Enumerator-Assessed Damage in Crop Cut Sub-Plot	0.005	0.008***	0.009***	0.010***	0.012***	0.016***	0.013***	0.015***	0.019***
Log Intercropping Seeding Rate (=100 for Pure stand Plots)	-0.073	-0.120	-0.016	-0.075	0.011	0.022	-0.028	-0.237	-0.696**
Cover Crops Present Prior to Planting†	-0.076	-0.114	-0.241	-0.205	-0.364**	-0.376	-0.143	0.073	0.021
Log Maize Seed Planting Rate (Kg/Ha)	0.054	0.122	0.207**	0.187***	0.293***	0.306**	0.349***	0.289**	0.393**
Organic Fertilizer Application†	0.283	-0.299	-0.110	-0.588	-0.745*	-0.629	-0.264	-0.372	-0.940
Inorganic Fertilizer Application†	-0.006	0.015	-0.045	-0.027	0.150	0.075	-0.131	0.009	0.087
Log Household Labor Days	0.004	0.017	0.053	0.093	0.094	0.138	0.067	0.146	0.412***
Log Hired Labor Days	0.128	0.083	0.075	0.165**	0.157**	0.046	0.104	0.211*	0.289*
No Hired Labor†	0.344	0.191	0.088	0.479*	0.508	0.140	0.230	0.166	0.103
% Deviation of Seasonal (May-June) Rainfall from Long-Term Mean	0.005	-0.021	0.021	0.029*	0.040**	0.060**	0.035	0.035	-0.032
Harvest Reported in More Than One Condition†	0.861***	0.406***	0.326**	0.292***	0.259**	0.390**	0.296*	0.219	0.204
<u>Household Characteristics</u>									
Wealth Index	-0.013	0.060	0.087	0.082	0.019	0.051	0.015	0.050	0.090
Agricultural Asset Index	0.029	0.095	0.074	0.080	0.038	-0.048	-0.010	0.022	0.137
Dependency Ratio	0.030	0.034	0.026	-0.008	0.019	-0.064	-0.083	-0.136	-0.164
Log Household Size	0.337**	0.034	0.036	0.002	0.066	0.086	0.073	0.042	-0.196
HH Cultivates More Than One Maize Plot	-0.001	-0.033	-0.088	0.033	0.028	0.017	-0.149	0.146	-0.059
<u>Manager Characteristics</u>									
Manager = Respondent†	0.028	0.001	0.014	0.032	0.164	0.388*	0.081	0.006	0.066
Received Crop-Production Related Extension Services†	-0.314*	-0.122	-0.123	0.015	0.055	0.138	0.145	-0.066	0.223
Female†	-0.178	-0.105	-0.114	0.037	-0.130	-0.049	-0.003	0.049	-0.105
Log Age (Years)	-0.010	0.027	0.229	0.124	0.344**	0.261	0.062	0.125	0.287
Log Years of Education	-0.004	0.023	0.057	0.019	0.030	0.012	-0.042	-0.139	-0.097
Round II Indicator (=1) †	0.542	-0.606	1.116	1.331**	1.827**	2.589**	1.144	1.471	-1.508
Constant	-2.959*	-0.875	-4.370***	-4.036***	-6.088***	-7.123***	-3.791**	-3.225	0.204
Enumerator Dummies?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Random Effects?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	474	474	474	474	474	474	474	474	474
R ²	0.144	0.153	0.188	0.206	0.271	0.242	0.23	0.252	0.257

Notes: ‡ Sample excludes plots that were subject to full-plot crop cut in MAPS II. † denotes a dummy variable. ***/**/* denote statistical significance at the 1/5/10 percent level, respectively. Production indicated as "heaped" if at least one condition of harvest was reported as 100, 200, 300, 400, or 500 kg, or 1, 2, 3, 4, 5, or 10 100 kg sacks. Overall R² reported for panel specifications.

Figures

Figure 1. Maize Area Harvested ('000s Ha), Uganda
FAOSTAT Annual Time Series

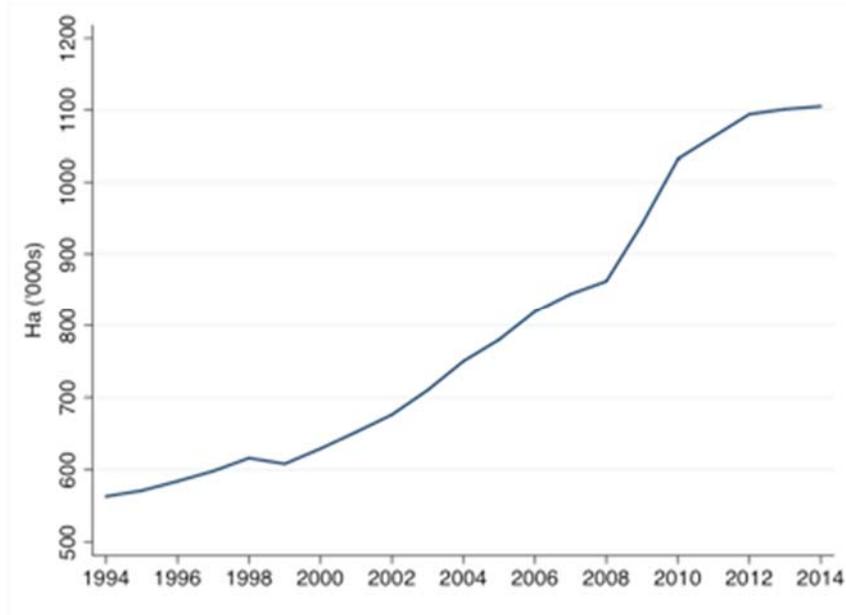


Figure 2. Maize Yield (Ton/Ha), Uganda
FAOSTAT Annual Time Series

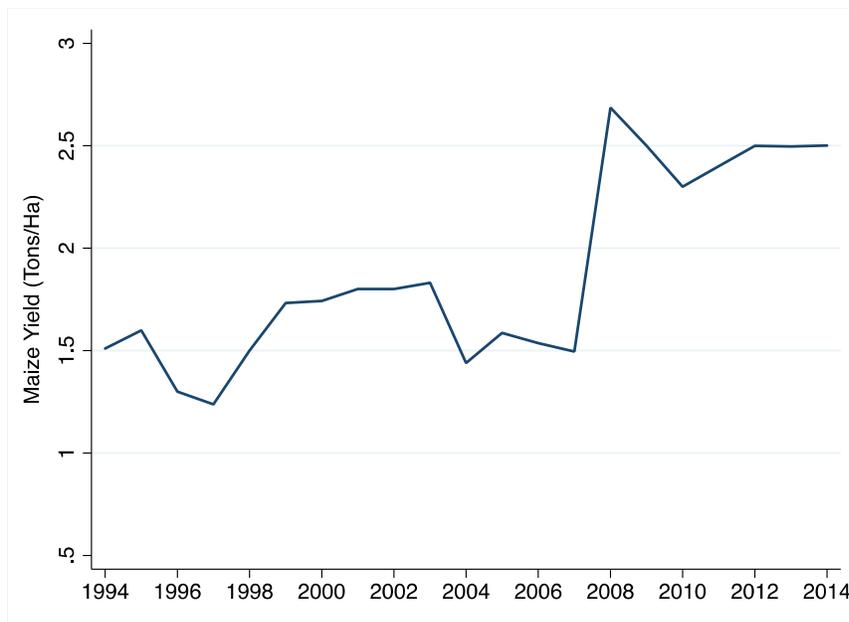
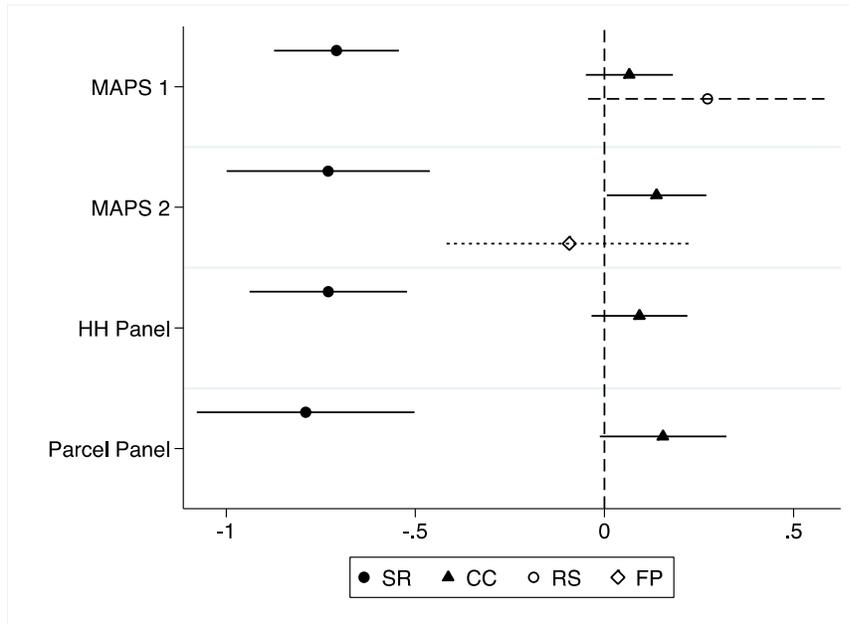
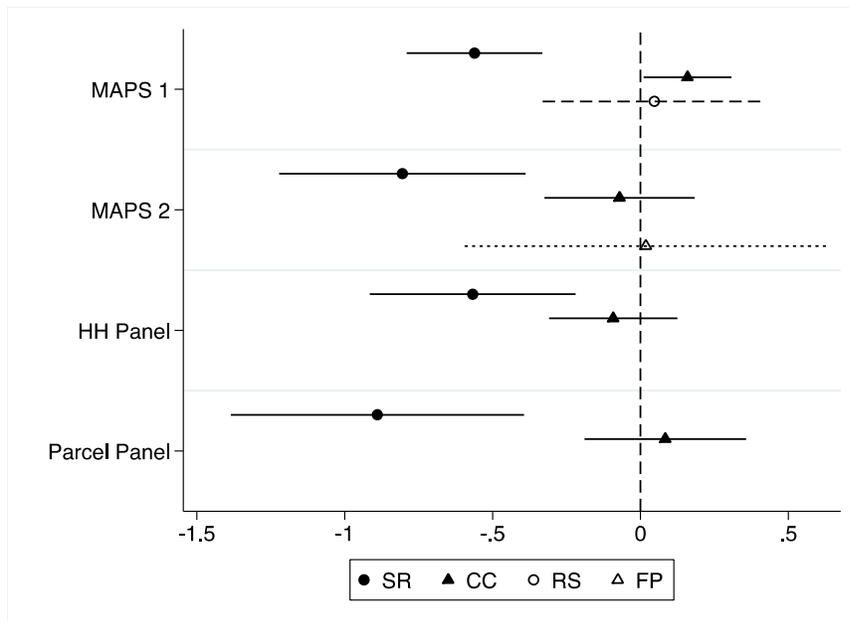


Figure 3. Selected Plot Area Coefficients w/ 95% Confidence Intervals - All Plots



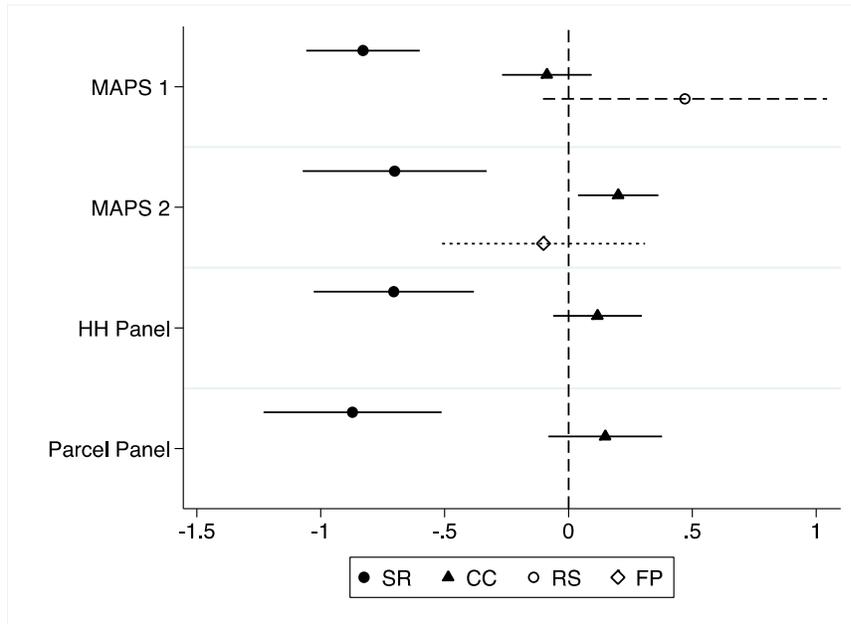
Notes: SR, CC, RS, and FP stand for self-reporting, sub-plot crop cutting, remote sensing, and full-plot crop cutting, respectively. MAPS II SR estimates are based on the plot sample net of those subject to full-plot crop cutting. Household and Parcel Panel are defined as in Table A4.1.

Figure 4. Selected Plot Area Coefficients w/ 95% Confidence Intervals – Pure Stand Plots



Notes: SR, CC, RS, and FP stand for self-reporting, sub-plot crop cutting, remote sensing, and full-plot crop cutting, respectively. MAPS II SR estimates are based on the plot sample net of those subject to full-plot crop cutting. Household and Parcel Panel are defined as in Table A4.2.

Figure 5 - Selected Plot Area Coefficients w/ 95% Confidence Intervals - Intercropped Plots



Notes: SR, CC, RS, and FP stand for self-reporting, sub-plot crop cutting, remote sensing, and full-plot crop cutting, respectively. MAPS II SR estimates are based on the plot sample net of those subject to full-plot crop cutting. Household and Parcel Panel are defined as in Table A4.3.

Figure 6. Consistency of Bias Across MAPS I and II

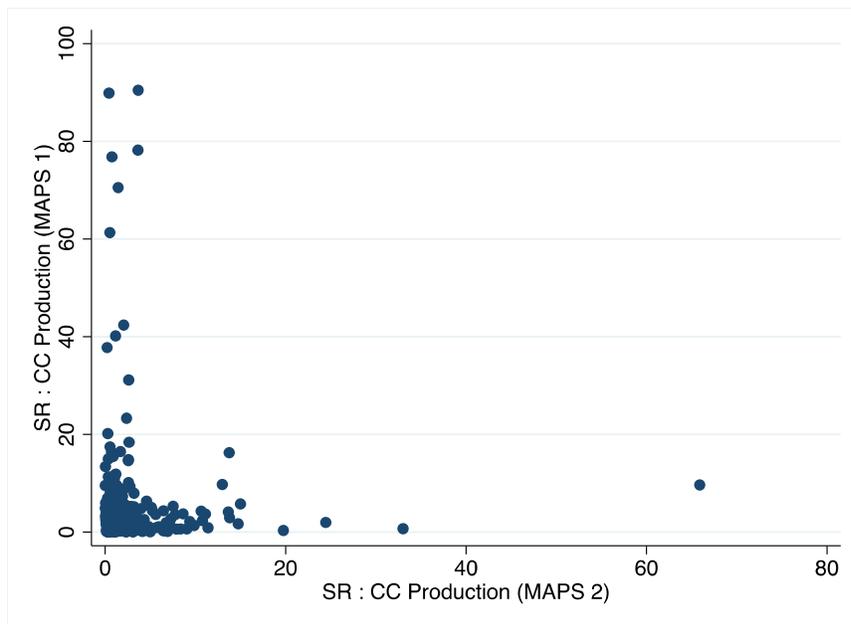


Figure 7. Over-Estimation in Self-Reported Yields by Round & GPS-Based Plot Area Quintile

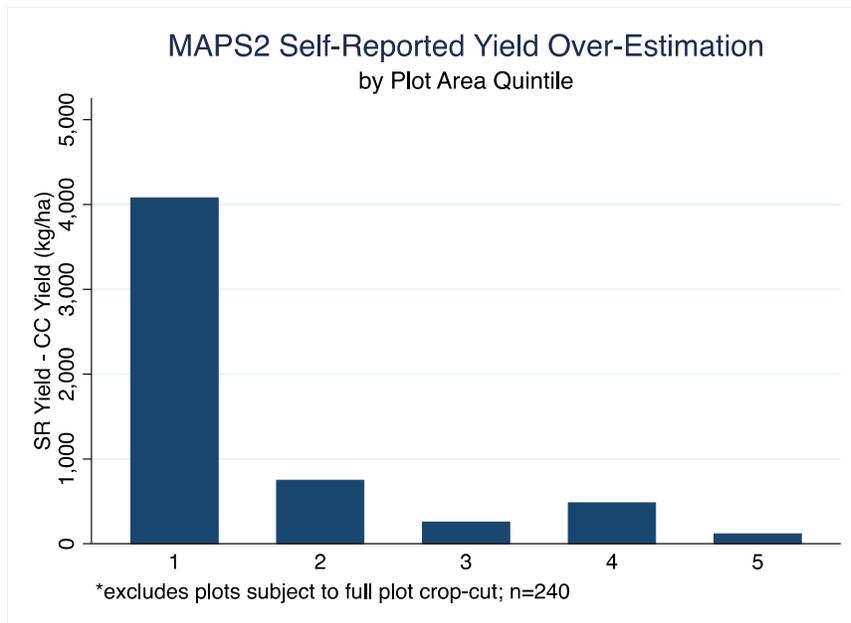
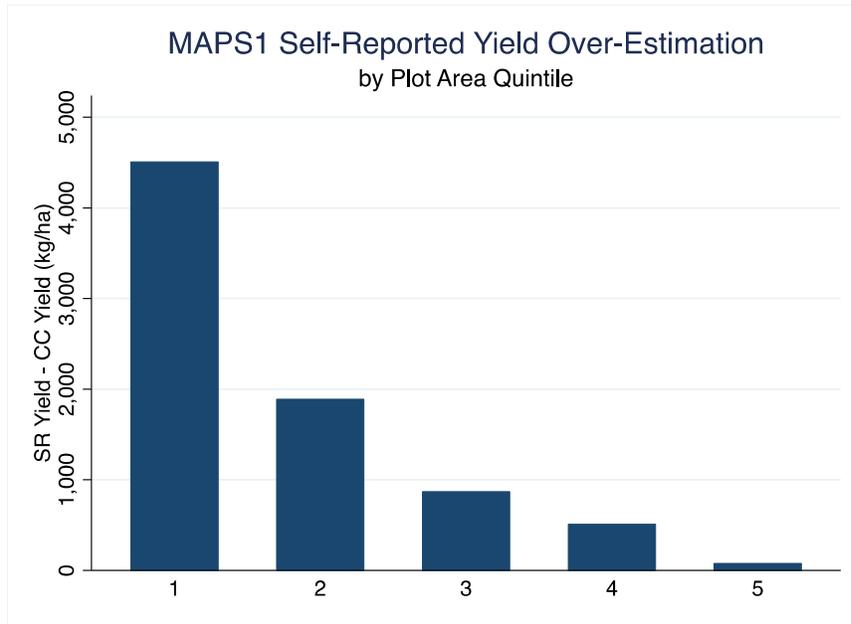
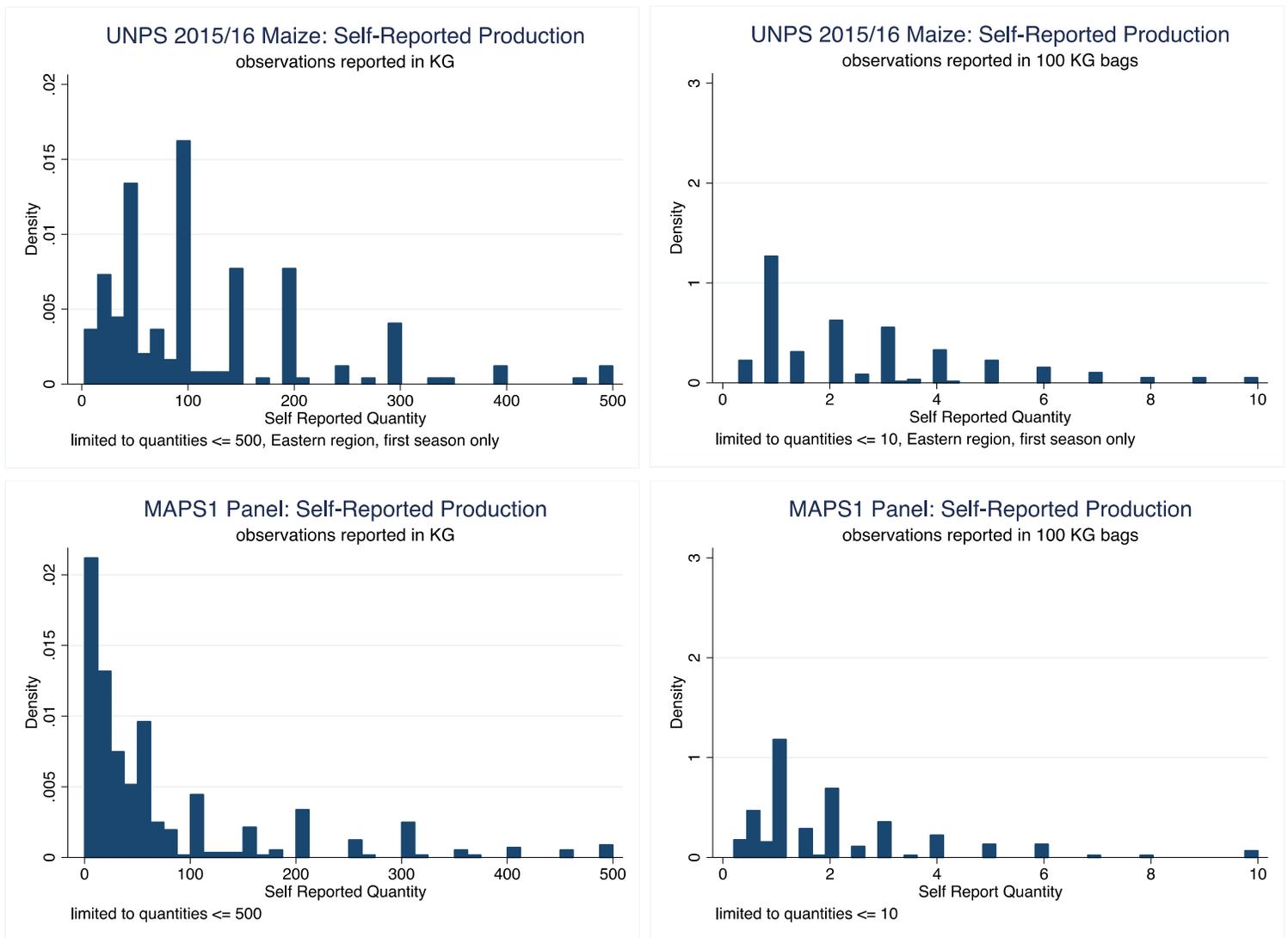


Figure 8. Comparisons of Heaping in Self-Reported Production Across UNPS 2015/16, MAPS I and MAPS II



Appendix Tables

Table A1. Variable Means by Round

	MAPS I	MAPS II	Difference		UNPS 2015-16 Eastern Region	UNPS 2015-16 MAPS Districts
Plot Area (GPS, ha)	0.14	0.18	0.03	***	-	-
Pure stand†	0.46	0.27	-0.19	***	0.30	0.19
Maize Seed Planted (Kg)	6.33	5.99	-0.33		6.59	5.72
Intercropping Seeding Rate (= 1 for Pure stand Plots)	0.83	0.81	-0.02		0.74	0.71
Organic Fertilizer Application†	0.01	0.02	0.01		0.02	0.01
Inorganic Fertilizer Application†	0.10	0.08	-0.02		0.04	0.05
Household Labor Days	50.09	49.50	-0.59		48.53	54.67
Hired Labor Days	5.50	5.68	0.18		1.40	0.95
No Hired Labor†	0.56	0.58	0.02		0.80	0.83
Female Plot Manager	0.40	0.49	0.09	***	0.34	0.25
Plot Manager Years of Education	6.32	6.34	0.02		7.70	7.98
Household Size	6.39	6.69	0.30		6.77	7.54
Dependency Ratio	1.44	1.37	-0.07		1.41	1.63
Incidence of HH level maize sales‡	N/A	N/A	N/A		41%	37%
Observations	440	440	440		457	81

Notes: † denotes a dummy variable. ‡ Incidence of maize sales of any quantity, of maize-growing households. Sales data not available in MAPS I or MAPS II due to timing of survey. *** denotes statistical significance of the mean difference at 1 percent level.

Table A2.1 Farmer-Reported Incidence of Damage on Selected Plot

MAPS I		MAPS II	
Any damage occurred	84%	Any damage occurred	96%
<i>Damage due to...</i>		<i>Damage due to...</i>	
Too much rain	16%	Unpredictable rain	6%
Too little rain	11%	Drought	37%
Locust	0%	Locust	0%
Termites	51%	Termites	44%
Other insects	28%	Other insects	16%
Crop disease	25%	Crop disease	20%
Weeds	24%	Weeds	21%
Hail	1%	Hail	0%
Floods	3%	Floods	0%
Frost	0%	Frost	0%
Wild Animals	3%	Wild Animals	5%
Domestic Animals	13%	Domestic Animals	12%
Birds	8%	Birds	4%
Shortage of seed	1%	Shortage of seed	0%
Bad soil	19%	Bad soil	15%
Security / theft	11%	Security / theft	16%
Spoiled seed	4%	Spoiled seed	1%
Observations	540	Observations	489

Table A2.2 Incidence of Damage & Pre-Harvest in Crop Cut Sub-Plot

	MAPS I	MAPS II
Incidence of Crop Cut Sub-Plot Damage	36%	89%
Mean % Damaged	29%	30%
Incidence of Crop Cut Sub-Plot Pre-Harvesting	8%	20%
Mean % Pre-Harvested	29%	26%

Table A3. Unconditional Quantile Regression (RIF): Log Plot-Level Maize Yield (Kg/Ha, GPS) - Household Panel Sample, All Plots

Dependent Variable	Self-Reported Yield‡								
	10	20	30	40	50	60	70	80	90
<i>Decile</i>									
<u>Plot Characteristics</u>									
Log Plot Area (GPS, ha)	-0.127	-0.312	-0.366**	-0.741***	-0.560***	-0.590***	-0.577***	-0.872***	-1.840***
Log Plot Distance from Dwelling (GPS, km)	0.395	0.206	0.132	0.258	0.147	0.003	-0.064	-0.135	0.144
Pure Stand†	0.546	0.599	0.048	-0.133	-0.342	-0.275	0.016	0.273	0.879
% Enumerator-Assessed Damage in Crop Cut									
Sub-Plot	-0.025**	-0.016	-0.012*	-0.013	-0.009	-0.007	-0.005	-0.003	0.001
Log Intercropping Seeding Rate (=100 for Pure Stand Plots)	-0.697*	-0.096	-0.123	-0.098	0.109	-0.073	0.066	-0.246	-0.748
Cover Crops Present Prior to Planting†	0.233	0.724	0.483	0.361	0.034	-0.175	0.045	0.056	0.434
Log Maize Seed Planted (Kg)	0.050	0.005	0.0800	0.185	0.142	0.026	0.086	0.260	0.519
Inorganic Fertilizer Application†	0.741	0.830	0.251	0.202	0.173	0.429	-0.002	-0.077	-0.261
Log Household Labor Days	0.367*	0.334	0.138	0.366*	0.184	0.373**	0.292**	0.318	0.453
Log Hired Labor Days	0.182	0.599**	0.253	0.373	0.244	0.297*	0.155	0.083	-0.140
No Hired Labor†	0.304	1.299**	-0.011	-0.228	0.101	0.101	0.003	-0.197	-0.663
% Deviation of Seasonal (May-June) Rainfall from Long-Term Mean	-0.024	-0.060	-0.005	-0.053	-0.003	0.006	0.018	-0.014	0.085
% of Crop Cut Sub-Plot within 4m of Plot Edge									
<u>Household Characteristics</u>									
Wealth Index	-0.251	0.075	0.065	0.116	0.084	0.156	0.090	0.092	-0.074
Agricultural Asset Index	-0.228	-0.230	-0.023	0.039	0.069	-0.064	-0.050	0.102	0.328
Dependency Ratio	0.172	0.171	0.000	0.007	0.207	0.155	0.018	-0.137	-0.159
Log Household Size	-0.043	0.035	-0.061	-0.692	-0.330	-0.201	0.589	0.284	0.504
<u>Manager Characteristics</u>									
Manager = Respondent†	0.679	0.535	0.254	0.481	0.272	0.355	0.180	0.503	0.779
Received Crop-Production Related Extension Services†	-0.199	-0.595	-0.291	-0.431	-0.114	-0.255	0.057	-0.126	-0.278
Female†	-0.067	-0.144	-0.235	0.154	0.276	0.286	0.060	0.366	-0.020
Log Age (Years)	0.966	0.720	0.235	0.272	0.128	-0.022	0.049	-0.321	-0.080
Log Years of Education	-0.361	-0.369	-0.295	-0.036	0.123	0.126	0.107	0.411	0.586
Round II Indicator (=1)†	-0.475	-2.237	-0.201	-2.335	-0.221	0.22	0.668	-0.686	3.084
Constant	4.756	3.723	5.634	6.995	4.164	4.209	2.14	5.536	-0.114
Household Fixed Effects?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	480	480	480	480	480	480	480	480	480
R²	0.179	0.178	0.182	0.253	0.238	0.247	0.271	0.289	0.304

Continued next page

Table A3. Unconditional Quantile Regression (RIF): Log Plot-Level Maize Yield (Kg/Ha, GPS) - Household Panel Sample, All Plots (Continued)

Dependent Variable	Sub-Plot Crop Cut Yield‡								
	10	20	30	40	50	60	70	80	90
<u>Plot Characteristics</u>									
Log Plot Area (GPS, ha)	-0.046	0.078	0.025	0.040	0.158	0.094	0.008	-0.023	0.196
Log Plot Distance from Dwelling (GPS, km)	0.165	0.316	0.273*	0.217**	0.252*	0.110	0.018	-0.053	-0.261
Pure Stand†	0.173	-0.387	-0.251	0.007	0.160	0.253	0.190	0.044	0.073
% Enumerator-Assessed Damage in Crop Cut Sub-Plot	-0.062***	-0.043***	-0.022***	-0.016***	-0.020***	-0.018***	-0.010***	-0.009***	-0.008*
Log Intercropping Seeding Rate (=100 for Pure Stand Plots)	0.193	0.401	0.250	-0.041	-0.057	-0.194	-0.149	-0.143	-0.204
Cover Crops Present Prior to Planting†	0.540	0.376	0.385	0.300	0.276	0.255	0.178	0.156	-0.012
Log Maize Seed Planted (Kg)	-0.414	-0.296	-0.189	0.006	-0.053	0.025	-0.038	-0.042	-0.059
Inorganic Fertilizer Application†	0.189	0.808	0.360	0.205	0.229	0.380	0.112	0.010	0.044
Log Household Labor Days	-0.077	0.015	-0.035	-0.112	-0.052	-0.041	-0.055	0.010	-0.175
Log Hired Labor Days	0.117	0.195	-0.048	-0.042	-0.036	-0.066	0.018	-0.025	0.045
No Hired Labor†	-0.139	0.424	-0.196	-0.081	-0.096	-0.314	-0.324	-0.349	0.174
% Deviation of Seasonal (May-June) Rainfall from Long-Term Mean	-0.040	-0.030	-0.021	-0.056*	-0.060	-0.078*	-0.066*	-0.078**	-0.076
% of Crop Cut Sub-Plot within 4m of Plot Edge	-0.934	-0.307	-0.270	-0.190	0.193	-0.026	0.051	0.039	0.064
<u>Household Characteristics</u>									
Wealth Index	0.218	0.167	0.102	-0.034	0.036	0.048	0.020	0.025	0.257
Agricultural Asset Index	-0.411	-0.233	-0.055	-0.100	-0.147	-0.170	-0.130	-0.076	-0.122
Dependency Ratio	-0.169	-0.299	-0.052	0.015	0.056	0.074	0.044	0.109	0.223
Log Household Size	-1.652	0.175	-0.378	-0.365	-0.053	-0.553	-0.568	-0.910**	-0.841
<u>Manager Characteristics</u>									
Manager = Respondent†	-0.222	0.224	0.051	0.228	0.233	0.182	0.024	-0.133	0.114
Received Crop-Production Related Extension Services†	0.151	-0.041	0.011	0.010	-0.046	-0.052	-0.050	-0.171	-0.137
Female†	0.319	0.167	0.072	0.041	-0.133	-0.038	-0.055	-0.173	-0.082
Log Age (Years)	1.593	0.474	-0.158	-0.181	-0.237	-0.118	-0.487	-0.451	0.215
Log Years of Education	-0.294	0.132	0.176	-0.068	-0.025	-0.064	-0.024	-0.036	-0.126
Round II Indicator (=1)†	-1.573	-1.362	-0.998	-2.360*	-2.385	-3.039*	-2.584*	-3.131**	-3.304
Constant	6.074	5.136	8.716***	11.696***	11.838***	13.467***	13.989***	15.029***	13.044***
Household Fixed Effects?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	480	480	480	480	480	480	480	480	480
R²	0.258	0.325	0.335	0.3	0.288	0.254	0.202	0.213	0.228

Continued next page

Table A3. Unconditional Quantile Regression (RIF): Log Plot-Level Maize Yield (Kg/Ha, GPS) - Household Panel Sample, All Plots (Continued)

Dependent Variable	Sub-Plot Crop Cut Yield (All Plots)								
	10	20	30	40	50	60	70	80	90
<u>Plot Characteristics</u>									
Log Plot Area (GPS, ha)	0.101	0.038	0.080	0.059	0.145	0.057	0.033	0.079	0.114
Log Plot Distance from Dwelling (GPS, km)	0.430	0.278*	0.075	0.066	0.087	0.024	-0.016	-0.041	-0.093
Pure Stand†	0.183	-0.036	-0.066	0.065	0.084	0.155	0.083	0.157	0.110
% Enumerator-Assessed Damage in Crop Cut Sub-Plot	-	-	-	-	-	-	-	-	-
Log Intercropping Seeding Rate (=100 for Pure Stand Plots)	0.065***	0.041***	0.025***	0.014***	-0.017***	0.013***	-0.008***	-0.007***	-0.005**
Cover Crops Present Prior to Planting†	0.337	0.285	0.157	0.067	0.035	-0.015	-0.039	-0.044	-0.010
Log Maize Seed Planted (Kg)	0.348	0.194	0.179	0.085	0.199	0.061	-0.046	-0.085	-0.066
Inorganic Fertilizer Application†	0.018	-0.005	0.017	-0.025	-0.068	-0.052	-0.036	-0.075	-0.156
Log Household Labor Days	0.056	0.736*	0.533*	0.233	0.407	0.291	0.207	0.141	0.151
Log Hired Labor Days	-0.235	0.010	-0.094	-0.005	0.046	0.004	0.020	0.029	-0.080
No Hired Labor†	0.128	-0.074	-0.188	-0.024	-0.101	0.020	0.059	0.100	0.143
% Deviation of Seasonal (May-June) Rainfall from Long-Term Mean	-0.114	-0.156	-0.435	-0.161	-0.308	-0.094	-0.130	0.045	0.212
% of Crop Cut Sub-Plot within 4m of Plot Edge	-0.049	-0.021	0.005	-0.006	-0.007	-0.013	-0.045*	-0.035	-0.020
<u>Household Characteristics</u>									
Wealth Index	-0.134	-0.220	-0.111	-0.088	0.179	0.143	0.112	0.070	0.029
Agricultural Asset Index	0.105	0.056	0.058	0.035	0.118	0.081	0.042	-0.012	0.115
Dependency Ratio	-0.264	-0.094	-0.037	-0.019	-0.049	-0.066	0.040	0.036	0.044
Log Household Size	-0.143	-0.091	-0.008	0.049	0.044	0.055	-0.010	0.107	0.108
<u>Manager Characteristics</u>									
Manager = Respondent†	-0.323	0.630	0.390	0.182	0.083	-0.004	-0.213	-0.285	-0.302
Received Crop-Production Related Extension Services†	0.100	0.158	0.004	0.055	-0.001	-0.057	-0.093	-0.150	-0.047
Female†	0.578	0.243	0.062	0.034	0.082	-0.022	-0.047	-0.111	-0.045
Log Age (Years)	0.030	-0.322	-0.201	-0.191	-0.366	-0.193	-0.179	-0.195	-0.072
Log Years of Education	0.331	-0.720	-0.695	-0.377	-0.716	-0.347	-0.422	-0.194	0.123
Round II Indicator (=1)†	0.061	0.242	0.178	0.001	0.053	0.001	0.039	-0.053	-0.174
Constant	-1.917	-0.866	0.056	-0.389	-0.312	-0.577	-1.951**	-1.553*	-0.972
Household Fixed Effects?	8.117	8.313**	8.440***	8.143***	10.124***	9.237***	11.544***	10.581***	9.206***
N	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R ²	880	880	880	880	880	880	880	880	880
	0.246	0.287	0.273	0.243	0.23	0.19	0.169	0.178	0.150

Notes: ‡ Sample excludes plots that were subject to full-plot crop cut in MAPS II. † denotes a dummy variable. ***/**/* denote statistical significance at the 1/5/10 percent level, respectively.

Table A4.1. OLS Regression Results - Dependent Variable: Log Plot-Level Maize Yield (Kg/Ha, GPS) - All Plots

Dependent Variable	MAPS I - Cross-Sectional						MAPS II - Cross-Sectional			
	Self-Reported Yield		Sub-Plot Crop Cut Yield		Remote Sensing Yield		Self-Reported Yield ‡	Sub-Plot Crop Cut Yield ‡	Sub-Plot Crop Cut Yield (All Plots)	Full Plot Crop Cut Yield
<i>Specification</i>	<i>1</i>	<i>2</i>	<i>3</i>	<i>4</i>	<i>5</i>	<i>6</i>	<i>7</i>	<i>8</i>	<i>9</i>	<i>10</i>
<u>Plot Characteristics</u>										
Log Plot Area (GPS, ha)	-0.712***	-0.708***	0.070	0.066	0.270*	0.273*	-0.731***	0.161*	0.138**	-0.093
Log Plot Distance from Dwelling (GPS, km)	-0.016	0.004	-0.080*	-0.061	-0.504***	-0.486***	0.043	0.087	0.004	-0.129
Pure Stand†	0.293	0.275	0.165*	0.144	-0.247	-0.264	0.255	0.239*	0.145	0.497***
% Enumerator-Assessed Damage in Crop Cut Sub-Plot	-0.010***	-0.009**	-0.020***	-0.018***	-0.001	0.000	-0.012***	-0.032***	-0.032***	-0.025***
Log Intercropping Seeding Rate (=100 for Pure Stand Plots)	-0.121	-0.085	0.167	0.202	-0.009	0.023	0.082	0.151	0.170***	0.294**
Cover Crops Present Prior to Planting†	0.351*	0.307*	0.190*	0.143	-0.081	-0.121	0.143	0.140	0.080	0.275
Log Maize Seed Planted (Kg)	0.190*	0.173*	-0.007	-0.024	-0.131	-0.146	0.396***	-0.053	-0.023	0.052
Inorganic Fertilizer Application†	0.249	0.301	0.127	0.189*	-0.549	-0.500	-0.133	-0.113	0.034	0.430**
Log Household Labor Days	0.380***	0.381***	0.020	0.015	0.172	0.171	0.364***	0.019	-0.019	-0.004
Log Hired Labor Days	0.287***	0.286***	0.001	0.004	0.013	0.013	0.162	-0.011	-0.033	-0.053
No Hired Labor†	-0.100	-0.116	-0.135	-0.146	0.221	0.207	-0.043	0.121	-0.086	-0.149
% Deviation of Seasonal (May-June) Rainfall from Long-Term Mean	-0.010	-0.004	-0.026***	-0.018**	0.063**	0.068**	-0.047	-0.001	0.030*	0.007
<u>Objectively-Measured Plot Characteristics</u>										
Soil Quality Index		1.262**		1.188***		1.113				
Genetic Heterogeneity of Plot's Maize		0.006		0.011***		0.007				
% of Crop Cut Sub-Plot within 4m of Plot Edge			0.014	-0.011				0.030	-0.001	
<u>Household Characteristics</u>										
Wealth Index	0.106	0.122	-0.009	0.007	-0.170	-0.155	0.138*	0.082	0.009	-0.079
Agricultural Asset Index	0.018	0.024	0.119***	0.128***	0.203	0.209	-0.032	-0.088	0.001	0.124
Dependency Ratio	-0.004	-0.016	0.090***	0.078**	-0.260*	-0.271*	0.003	-0.041	0.036	0.070
Log Household Size	-0.114	-0.103	-0.014	-0.003	0.149	0.159	-0.131	-0.027	-0.043	0.044
<u>Manager Characteristics</u>										
Manager = Respondent†	0.121	0.150	0.036	0.059	0.193	0.218	0.121	0.036	-0.020	-0.143
Received Crop-Production Related Extension Services†	0.068	0.077	-0.126	-0.129	0.023	0.028	-0.206	-0.138	0.006	-0.069
Female†	-0.058	-0.09	0.035	-0.011	-0.074	-0.106	0.018	-0.103	-0.076	-0.262

Continued next page

Table A4.1. OLS Regression Results - Dependent Variable: Log Plot-Level Maize Yield (Kg/Ha, GPS) - All Plots (continued)

Dependent Variable	MAPS I - Cross-Sectional				MAPS II - Cross-Sectional					
	Self-Reported Yield		Sub-Plot Crop Cut Yield		Remote Sensing Yield		Self-Reported Yield ‡	Sub-Plot Crop Cut Yield ‡	Sub-Plot Crop Cut Yield (All Plots)	Full Plot Crop Cut Yield
<i>Specification</i>	<i>1</i>	<i>2</i>	<i>3</i>	<i>4</i>	<i>5</i>	<i>6</i>	<i>7</i>	<i>8</i>	<i>9</i>	<i>10</i>
Log Age (Years)	-0.112	-0.109	-0.024	-0.019	-0.075	-0.071	-0.074	0.085	0.166	0.407*
Log Years of Education	-0.074	-0.086	-0.010	-0.020	0.330*	0.320*	0.127	0.110	0.058	0.123
Round II Indicator (=1)†										
Constant	5.003***	3.450*	7.285***	5.369***	2.204	0.73	3.704***	6.478***	5.944***	3.207**
Fixed Effects?	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
N	440	440	440	440	440	440	240	240	440	200
R²	0.233	0.245	0.338	0.384	0.092	0.096	0.249	0.465	0.477	0.361

Continued next page

Table A4.1. OLS Regression Results - Dependent Variable: Log Plot-Level Maize Yield (Kg/Ha, GPS) - All Plots (continued)

Dependent Variable	Household Panel			Parcel Panel		
	Self-Reported Yield ‡	Sub-Plot Crop Cut Yield ‡	Sub-Plot Crop Cut Yield (All Plots)	Self-Reported Yield ‡	Sub-Plot Crop Cut Yield ‡	Sub-Plot Crop Cut Yield (All Plots)
<i>Specification</i>	<i>11</i>	<i>12</i>	<i>13</i>	<i>14</i>	<i>15</i>	<i>16</i>
<u>Plot Characteristics</u>						
Log Plot Area (GPS, ha)	-0.730***	0.063	0.093	-0.790***	0.116	0.155*
Log Plot Distance from Dwelling (GPS, km)	0.251**	0.118	0.077	0.278	0.113	0.099*
Pure Stand†	0.085	0.116	0.119	0.193	0.132	0.204
% Enumerator-Assessed Damage in Crop Cut Sub-Plot	-0.015***	-0.028***	-0.024***	-0.009*	-0.025***	-0.024***
Log Intercropping Seeding Rate (=100 for Pure Stand Plots)	-0.170	0.058	0.121	-0.238	-0.036	0.088
Cover Crops Present Prior to Planting†	0.178	0.353***	0.141	0.034	0.353*	0.179
Log Maize Seed Planted (Kg)	0.135	-0.053	-0.027	0.206	-0.037	-0.010
Inorganic Fertilizer Application†	0.224	0.377*	0.437**	0.372	0.374	0.500**
Log Household Labor Days	0.401***	-0.091	-0.085	0.436***	-0.062	-0.096
Log Hired Labor Days	0.257**	-0.039	-0.002	0.179	0.026	0.096
No Hired Labor†	0.229	-0.138	-0.110	0.374	0.003	0.118
% Deviation of Seasonal (May-June) Rainfall from Long-Term Mean	-0.015	-0.050**	-0.023	-0.028	-0.034	-0.020
<u>Objectively-Measured Plot Characteristics</u>						
Soil Quality Index						
Genetic Heterogeneity of Plot's Maize						
% of Crop Cut Sub-Plot within 4m of Plot Edge		-0.011	0.089		-0.060	0.034
<u>Household Characteristics</u>						
Wealth Index	0.022	0.167	0.116*	-0.024	0.150	0.125
Agricultural Asset Index	-0.039	-0.171**	-0.064	-0.081	-0.152*	-0.058
Dependency Ratio	0.044	-0.070	-0.041	-0.060	-0.091	-0.022
Log Household Size	0.006	-0.087	0.191	0.078	0.105	0.266
<u>Manager Characteristics</u>						
Manager = Respondent†	0.471**	0.058	-0.008	0.519**	0.059	-0.049
Received Crop-Production Related Extension Services†	-0.343*	0.043	0.064	-0.162	0.178	0.099
Female†	0.236	-0.189	-0.263**	-0.075	-0.125	-0.123
Log Age (Years)	0.713	-0.050	-0.269	-0.254	0.004	-0.245
Log Years of Education	-0.081	-0.154	-0.004	-0.066	-0.089	0.061

Continued next page

Table A4.1. OLS Regression Results - Dependent Variable: Log Plot-Level Maize Yield (Kg/Ha, GPS) - All Plots (continued)

Dependent Variable	Household Panel			Parcel Panel		
	Self-Reported Yield ‡	Sub-Plot Crop Cut Yield ‡	Sub-Plot Crop Cut Yield (All Plots)	Self-Reported Yield ‡	Sub-Plot Crop Cut Yield ‡	Sub-Plot Crop Cut Yield (All Plots)
<i>Specification</i>	<i>11</i>	<i>12</i>	<i>13</i>	<i>14</i>	<i>15</i>	<i>16</i>
Round II Indicator (=1)†	-0.545	-1.948**	-0.929	-1.309	-1.449	-0.906
Constant	2.479	10.403***	8.892***	6.590*	9.253***	8.553***
Fixed Effects?	Household	Household	Household	Parcel	Parcel	Parcel
N	480	480	880	316	316	594
R²	0.334	0.489	0.448	0.423	0.454	0.458

Notes: ‡ Sample excludes plots that were subject to full-plot crop cut in MAPS II. † denotes a dummy variable. ***/**/* denote statistical significance at the 1/5/10 percent level, respectively.

Table A4.2. OLS Regression Results - Dependent Variable: Log Plot-Level Maize Yield (Kg/Ha, GPS) - Pure Stand Plots

Dependent Variable	MAPS I - Cross-Sectional						MAPS II - Cross-Sectional			
	Self-Reported Yield		Sub-Plot Crop Cut Yield		Remote Sensing Yield		Self-Reported Yield ‡	Sub-Plot Crop Cut Yield ‡	Sub-Plot Crop Cut Yield (All Plots)	Full Plot Crop Cut Yield
<i>Specification</i>	<i>1</i>	<i>2</i>	<i>3</i>	<i>4</i>	<i>5</i>	<i>6</i>	<i>7</i>	<i>8</i>	<i>9</i>	<i>10</i>
<u>Plot Characteristics</u>										
Log Plot Area (GPS, ha)	-0.584***	-0.561***	0.161**	0.159**	0.024	0.047	-0.805***	-0.169	-0.071	0.018
Log Plot Distance from Dwelling (GPS, km)	-0.090	-0.063	-0.105*	-0.076	-0.378**	-0.350**	0.138	0.065	-0.037	-0.118
Pure Stand†	(omitted)	(omitted)	(omitted)	(omitted)	(omitted)	(omitted)	(omitted)	(omitted)	(omitted)	(omitted)
% Enumerator-Assessed Damage in Crop Cut Sub-Plot	-0.007	-0.005	-0.016***	-0.014***	-0.017	-0.016	-0.011*	-0.022***	-0.024***	-0.008
Log Intercropping Seeding Rate (=100 for Pure Stand Plots)	(omitted)	(omitted)	(omitted)	(omitted)	(omitted)	(omitted)	(omitted)	(omitted)	(omitted)	(omitted)
Cover Crops Present Prior to Planting†	-0.070	-0.083	0.032	-0.006	-0.173	-0.188	-0.038	0.163	0.026	0.548
Log Maize Seed Planted (Kg)	0.097	0.080	-0.006	-0.022	-0.261	-0.278	0.401**	0.169	0.203	0.158
Inorganic Fertilizer Application†	0.258	0.281	0.089	0.137	-0.215	-0.189	0.409	0.215	0.153	0.666*
Log Household Labor Days	0.445**	0.457**	-0.024	-0.026	0.014	0.026	0.425**	0.134**	0.030	-0.098
Log Hired Labor Days	0.317**	0.325**	-0.082	-0.06	0.014	0.023	0.095	-0.044	-0.010	-0.010
No Hired Labor†	-0.072	-0.087	-0.097	-0.081	-0.167	-0.182	-0.332	0.098	-0.081	0.333
% Deviation of Seasonal (May-June) Rainfall from Long-Term Mean	0.030	0.039*	-0.011	0.003	0.057	0.066	0.016	0.027	0.025	0.000
<u>Objectively-Measured Plot Characteristics</u>										
Soil Quality Index		1.519**		1.559***		1.615				
Genetic Heterogeneity of Plot's Maize		0.002		0.010**		0.003				
% of Crop Cut Sub-Plot within 4m of Plot Edge			-0.084	-0.202				-0.888**	-0.279	
<u>Household Characteristics</u>										
Wealth Index	0.082	0.100	0.029	0.055	0.058	0.078	0.199	0.055	-0.063	-0.224
Agricultural Asset Index	-0.076	-0.058	0.113*	0.144**	-0.070	-0.051	-0.043	-0.097	-0.005	0.136
Dependency Ratio	0.043	0.039	0.073	0.067	-0.355	-0.359	0.103	0.034	-0.032	0.032
Log Household Size	-0.239	-0.278	-0.108	-0.137	0.394	0.354	-0.156	-0.077	0.136	-0.098
<u>Manager Characteristics</u>										
Manager = Respondent†	-0.326	-0.271	-0.076	-0.017	0.625	0.684	-0.352	0.208	0.038	-0.354
Received Crop-Production Related Extension Services†	0.212	0.258	-0.093	-0.086	-0.316	-0.269	-0.278	-0.242	0.099	0.742
Female†	-0.223	-0.209	-0.105	-0.130	-0.616	-0.604	0.327	-0.268	-0.083	-0.073
Log Age (Years)	0.289	0.323	0.121	0.163	0.268	0.304	0.067	-0.388	-0.328	0.073

Continued next page

Table A4.2. OLS Regression Results - Dependent Variable: Log Plot-Level Maize Yield (Kg/Ha, GPS) - Pure Stand Plots (continued)

Dependent Variable	MAPS I - Cross-Sectional						MAPS II - Cross-Sectional			
	Self-Reported Yield		Sub-Plot Crop Cut Yield		Remote Sensing Yield		Self-Reported Yield ‡	Sub-Plot Crop Cut Yield ‡	Sub-Plot Crop Cut Yield (All Plots)	Full Plot Crop Cut Yield
<i>Specification</i>	<i>1</i>	<i>2</i>	<i>3</i>	<i>4</i>	<i>5</i>	<i>6</i>	<i>7</i>	<i>8</i>	<i>9</i>	<i>10</i>
Log Years of Education	-0.154	-0.160	-0.022	-0.031	0.459*	0.452*	0.228	0.001	0.008	0.230
Round II Indicator (=1)†										
Constant	2.037	0.518	7.705***	5.399***	0.990	-0.668	3.456	7.694***	7.284***	5.858**
Fixed Effects?	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
N	201	201	201	201	201	201	70	70	117	47
R²	0.233	0.247	0.271	0.351	0.128	0.135	0.349	0.632	0.463	0.416

Continued next page

Table A4.2. OLS Regression Results - Dependent Variable: Log Plot-Level Maize Yield (Kg/Ha, GPS) - Pure Stand Plots (continued)

Dependent Variable	Household Panel ^o			Parcel Panel ^o		
	Self-Reported Yield ‡	Sub-Plot Crop Cut Yield ‡	Sub-Plot Crop Cut Yield (All Plots)	Self-Reported Yield ‡	Sub-Plot Crop Cut Yield ‡	Sub-Plot Crop Cut Yield (All Plots)
<i>Specification</i>	<i>11</i>	<i>12</i>	<i>13</i>	<i>14</i>	<i>15</i>	<i>16</i>
<u>Plot Characteristics</u>						
Log Plot Area (GPS, ha)	-0.346*	0.084	-0.092	-0.890***	-0.124	-0.126
Log Plot Distance from Dwelling (GPS, km)	0.000	0.089	0.047	0.556	-0.162	0.037
Pure Stand†	(omitted)	(omitted)	(omitted)	(omitted)	(omitted)	(omitted)
% Enumerator-Assessed Damage in Crop Cut Sub-Plot	-0.010	-0.017**	-0.023***	-0.044***	-0.008	-0.021***
Log Intercropping Seeding Rate (=100 for Pure Stand Plots)	(omitted)	(omitted)	(omitted)	(omitted)	(omitted)	(omitted)
Cover Crops Present Prior to Planting†	-0.470	0.168	-0.155	-1.236**	-0.005	-0.075
Log Maize Seed Planted (Kg)	0.385	0.343*	0.323***	0.552**	0.131	0.337***
Inorganic Fertilizer Application†	-0.229	0.210	0.119	-0.345	0.207	0.174
Log Household Labor Days	0.332	-0.155	-0.054	0.601***	0.036	0.018
Log Hired Labor Days	0.343**	0.117	-0.125	-0.174	-0.163	-0.211**
No Hired Labor†	0.983**	0.916**	0.137	1.320*	0.496*	-0.043
% Deviation of Seasonal (May-June) Rainfall from Long-Term Mean	0.020	-0.044	-0.007	-0.014	-0.061	-0.054
<u>Objectively-Measured Plot Characteristics</u>						
Soil Quality Index						
Genetic Heterogeneity of Plot's Maize						
% of Crop Cut Sub-Plot within 4m of Plot Edge		0.435	0.096		-0.178	0.156
<u>Household Characteristics</u>						
Wealth Index	-0.174	0.217	0.037	0.317	0.223	0.172
Agricultural Asset Index	0.241	0.030	0.104	0.363*	0.025	0.003
Dependency Ratio	0.011	-0.101	-0.093	0.061	-0.169	-0.109
Log Household Size	-1.158	0.154	0.103	0.912	0.291	-0.355
<u>Manager Characteristics</u>						
Manager = Respondent†	-0.844***	-0.416*	-0.212	-0.352	-0.395**	-0.293
Received Crop-Production Related Extension Services†	-0.076	-0.012	-0.018	-0.716*	-0.259	-0.14
Female†	0.611*	-0.345	-0.393*	-1.093	-0.631**	-0.289
Log Age (Years)	-0.911	-1.298**	-0.455	-2.025***	-0.662	-0.306
Log Years of Education	0.477*	0.069	0.070	-0.351	0.128	-0.211
Round II Indicator (=1)†	0.371	-1.913	-0.231	0.292	-2.423	-2.033

Continued next page

Table A4.2. OLS Regression Results - Dependent Variable: Log Plot-Level Maize Yield (Kg/Ha, GPS) - Pure Stand Plots (continued)

Dependent Variable	Household Panel ^o			Parcel Panel ^o		
	Self-Reported Yield ‡	Sub-Plot Crop Cut Yield ‡	Sub-Plot Crop Cut Yield (All Plots)	Self-Reported Yield ‡	Sub-Plot Crop Cut Yield ‡	Sub-Plot Crop Cut Yield (All Plots)
<i>Specification</i>	<i>11</i>	<i>12</i>	<i>13</i>	<i>14</i>	<i>15</i>	<i>16</i>
Constant	8.671*	13.752***	8.801***	11.620*	11.459***	11.462***
Fixed Effects?	Household	Household	Household	Parcel	Parcel	Parcel
N	118	118	186	64	64	114
R²	0.514	0.592	0.527	0.822	0.837	0.655

Notes: ^oLimited to households with Pure Stand plots in both waves. ‡ Sample excludes plots that were subject to full-plot crop cut in MAPS II. † denotes a dummy variable. ***/**/* denote statistical significance at the 1/5/10 percent level, respectively.

Table A4.3. OLS Regression Results - Dependent Variable: Log Plot-Level Maize Yield (Kg/Ha, GPS) - Intercropped Plots

Dependent Variable	MAPS I - Cross-Sectional						MAPS II - Cross-Sectional			
	Self-Reported Yield		Sub-Plot Crop Cut Yield		Remote Sensing Yield		Self-Reported Yield ‡	Sub-Plot Crop Cut Yield ‡	Sub-Plot Crop Cut Yield (All Plots)	Full Plot Crop Cut Yield
<i>Specification</i>	<i>1</i>	<i>2</i>	<i>3</i>	<i>4</i>	<i>5</i>	<i>6</i>	<i>7</i>	<i>8</i>	<i>9</i>	<i>10</i>
<u>Plot Characteristics</u>										
Log Plot Area (GPS, ha)	-0.813***	-0.829***	-0.077	-0.087	0.479	0.470	-0.702***	0.260**	0.201**	-0.101
Log Plot Distance from Dwelling (GPS, km)	0.013	0.026	-0.093	-0.080	-0.602***	-0.602***	0.002	0.070	-0.002	-0.125
Pure Stand†	(omitted)	(omitted)	(omitted)	(omitted)	(omitted)	(omitted)	(omitted)	(omitted)	(omitted)	(omitted)
% Enumerator-Assessed Damage in Crop Cut Sub-Plot	-0.010**	-0.009*	-0.021***	-0.020***	0.010	0.011	-0.013***	-0.034***	-0.034***	-0.031***
Log Intercropping Seeding Rate (=100 for Pure Stand Plots)	-0.165	-0.122	0.196	0.241	-0.08	-0.067	0.070	0.181*	0.180***	0.303**
Intercropped with Legumes†	0.210	0.194	0.097	0.087	0.773*	0.794*	0.149	0.269	0.138	0.004
Cover Crops Present Prior to Planting†	0.474**	0.431*	0.242**	0.205*	-0.128	-0.113	0.131	0.168	0.077	0.293
Log Maize Seed Planted (Kg)	0.251**	0.227**	0.008	-0.020	0.046	0.034	0.384**	-0.128	-0.073	0.016
Inorganic Fertilizer Application†	0.019	0.046	0.106	0.132	-0.306	-0.315	-0.652	-0.126	-0.017	0.370*
Log Household Labor Days	0.294**	0.298**	0.055	0.057	0.259	0.250	0.290**	-0.029	-0.044	0.155
Log Hired Labor Days	0.290***	0.271***	0.074	0.056	0.035	0.028	0.159	-0.021	-0.082	-0.206
No Hired Labor†	-0.081	-0.133	-0.176	-0.231	0.535	0.500	-0.073	0.068	-0.119	-0.370
% Deviation of Seasonal (May-June) Rainfall from Long-Term Mean	-0.034**	-0.028	-0.036***	-0.028**	0.048	0.055	-0.065	0.007	0.032	-0.009
<u>Objectively-Measured Plot Characteristics</u>										
Soil Quality Index		1.097		0.986**		-0.290				
Genetic Heterogeneity of Plot's Maize		0.011		0.014**		0.015				
% of Crop Cut Sub-Plot within 4m of Plot Edge			-0.016	0.017				0.393	0.116	
<u>Household Characteristics</u>										
Wealth Index	0.104	0.114	-0.034	-0.024	-0.415*	-0.422*	0.140	0.092	0.032	-0.089
Agricultural Asset Index	0.066	0.062	0.140**	0.135**	0.462**	0.461**	-0.061	-0.082	0.003	0.178
Dependency Ratio	-0.039	-0.058	0.128**	0.108**	-0.250	-0.247	-0.008	-0.063	0.048	0.121
Log Household Size	0.028	0.042	0.038	0.044	0.045	0.017	-0.082	0.019	-0.076	0.046
<u>Manager Characteristics</u>										
Manager = Respondent†	0.646**	0.651***	0.081	0.079	0.046	0.007	0.379	0.048	0.031	0.146
Received Crop-Production Related Extension Services†	-0.027	-0.024	-0.099	-0.095	0.410	0.419	-0.231	-0.090	-0.025	-0.465**
Female†	0.036	-0.002	0.109	0.067	0.119	0.107	-0.104	-0.068	-0.118	-0.444**

Continued next page

Table A4.3. OLS Regression Results - Dependent Variable: Log Plot-Level Maize Yield (Kg/Ha, GPS) - Intercropped Plots (continued)

Dependent Variable	MAPS I - Cross-Sectional						MAPS II - Cross-Sectional			
	Self-Reported Yield		Sub-Plot Crop Cut Yield		Remote Sensing Yield		Self-Reported Yield ‡	Sub-Plot Crop Cut Yield ‡	Sub-Plot Crop Cut Yield (All Plots)	Full Plot Crop Cut Yield
<i>Specification</i>	<i>1</i>	<i>2</i>	<i>3</i>	<i>4</i>	<i>5</i>	<i>6</i>	<i>7</i>	<i>8</i>	<i>9</i>	<i>10</i>
Log Age (Years)	-0.387	-0.394*	-0.170	-0.174	-0.264	-0.261	-0.131	0.138	0.260**	0.386
Log Years of Education	-0.008	-0.016	-0.015	-0.017	0.050	0.067	0.053	0.109	0.073	0.178
Round II Indicator (=1)†										
Constant	6.463***	4.694*	7.354***	5.409***	3.312	2.256	4.274***	6.314***	5.864***	3.103**
Fixed Effects?	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
N	239	239	239	239	239	239	170	170	323	153
R²	0.292	0.306	0.396	0.436	0.156	0.160	0.243	0.473	0.499	0.403

Continued next page

Table A4.3. OLS Regression Results - Dependent Variable: Log Plot-Level Maize Yield (Kg/Ha, GPS) - Intercropped Plots (continued)

Dependent Variable	Household Panel ^o			Parcel Panel ^o		
	Self-Reported Yield ‡	Sub-Plot Crop Cut Yield ‡	Sub-Plot Crop Cut Yield (All Plots)	Self-Reported Yield ‡	Sub-Plot Crop Cut Yield ‡	Sub-Plot Crop Cut Yield (All Plots)
<i>Specification</i>	<i>11</i>	<i>12</i>	<i>13</i>	<i>14</i>	<i>15</i>	<i>16</i>
<u>Plot Characteristics</u>						
Log Plot Area (GPS, ha)	-0.705***	0.146	0.117	-0.872***	0.208	0.148
Log Plot Distance from Dwelling (GPS, km)	0.362*	0.091	0.043	0.019	0.104	0.073
Pure Stand†	(omitted)	(omitted)	(omitted)	(omitted)	(omitted)	(omitted)
% Enumerator-Assessed Damage in Crop Cut Sub-Plot	-0.016**	-0.031***	-0.026***	-0.006	-0.032***	-0.026***
Log Intercropping Seeding Rate (=100 for Pure Stand Plots)	-0.116	0.062	0.110	-0.134	0.060	0.119
Intercropped with Legumes†	0.434	0.046	0.008	0.524	0.204	0.135
Cover Crops Present Prior to Planting†	0.390	0.194	0.096	0.416	0.256	0.125
Log Maize Seed Planted (Kg)	0.012	-0.111	-0.072	0.031	-0.169	-0.036
Inorganic Fertilizer Application†	0.566	0.849*	0.549	1.217***	1.388	0.768
Log Household Labor Days	0.148	-0.148	0.016	0.275	-0.171	-0.041
Log Hired Labor Days	0.309*	0.075	0.014	0.183	0.115	0.023
No Hired Labor†	0.387	-0.035	-0.143	-0.026	0.030	-0.038
% Deviation of Seasonal (May-June) Rainfall from Long-Term Mean	-0.027	-0.015	-0.036*	-0.072	-0.007	-0.024
<u>Objectively-Measured Plot Characteristics</u>						
Soil Quality Index						
Genetic Heterogeneity of Plot's Maize						
% of Crop Cut Sub-Plot within 4m of Plot Edge		0.128	0.281		-0.070	0.120
<u>Household Characteristics</u>						
Wealth Index	-0.025	0.183	0.116	-0.148	0.147	0.109
Agricultural Asset Index	-0.079	-0.143	-0.102	-0.131	-0.192*	-0.168*
Dependency Ratio	-0.018	-0.144	-0.044	0.033	-0.009	0.045
Log Household Size	0.719	0.479	0.541	0.456	0.496	0.650
<u>Manager Characteristics</u>						
Manager = Respondent†	0.912***	0.246	0.107	0.849***	-0.075	-0.130
Received Crop-Production Related Extension Services†	-0.537*	0.143	0.144	-0.124	0.289	0.224
Female†	-0.192	0.112	-0.121	-0.050	0.333	-0.018
Log Age (Years)	-0.392	0.275	-0.308	-0.773*	0.429	-0.147
Log Years of Education	-0.018	-0.202	-0.135	0.159	-0.144	-0.120
Round II Indicator (=1)†	-0.914	-0.585	-1.425*	-2.907	-0.256	-0.964

Continued next page

Table A4.3. OLS Regression Results - Dependent Variable: Log Plot-Level Maize Yield (Kg/Ha, GPS) - Intercropped Plots (continued)

Dependent Variable	Household Panel ^o			Parcel Panel ^o		
	Self-Reported Yield ‡	Sub-Plot Crop Cut Yield ‡	Sub-Plot Crop Cut Yield (All Plots)	Self-Reported Yield ‡	Sub-Plot Crop Cut Yield ‡	Sub-Plot Crop Cut Yield (All Plots)
<i>Specification</i>	<i>11</i>	<i>12</i>	<i>13</i>	<i>14</i>	<i>15</i>	<i>16</i>
Constant	6.216	6.684**	8.730***	8.230*	5.711	7.494***
Fixed Effects?	Household	Household	Household	Parcel	Parcel	Parcel
N	240	240	430	170	170	312
R²	0.389	0.550	0.467	0.490	0.564	0.476

Notes: ^oLimited to households with intercropped plots in both waves. ‡ Sample excludes plots that were subject to full-plot crop cut in MAPS II. † denotes a dummy variable. ***/**/* denote statistical significance at the 1/5/10 percent level, respectively.

Appendix I - MAPS II Household Tracking Protocol

The following is an excerpt from the MAPS II field staff manual:

*“If (i) the **head of household** is in the MAPS I dwelling:*

- Consider the household to be tracked.
- Mark the ID of the head and all MAPS I members living with the head.

*If (i) the **head of household** has moved out of the MAPS I dwelling AND (ii) the head of household lives in the same enumeration area (EA) or in its close vicinity (in Iganga or Mayuge), either with the entire set or a subset of MAPS I household members or by himself:*

- Track the head living in the same EA or in its close vicinity. Once located, consider the household to be tracked.
- Mark the ID of the head and all MAPS I household members living with the head.

*If (i) the **head of household** has moved out of the MAPS I dwelling, AND (ii) the head of household DOES NOT live in the same EA or in its close vicinity (in Iganga or Mayuge), AND (iii) the spouse or other MAPS I household members live in the EA or in its close vicinity (in Iganga or Mayuge) - including the possibility of living in the MAPS I dwelling:*

Scenario #1:

- Attempt to track first the spouse living in the same EA or in its close vicinity. Once located, consider the household to be tracked.
- Mark the ID of the spouse and all MAPS I household members living with the spouse.

Scenario #2:

- If there was no spouse in MAPS I or the spouse is not in the same EA or in its close vicinity, locate the largest group of MAPS I household members living in the same EA or in its close vicinity. If there is a tie in group size, locate the group members that live closest to the MAPS I dwelling location. If there is yet another tie in terms of distance, locate the group that contains the lowest MAPS I PID. Once the group is located, consider the household to be tracked.
- Mark the ID of the MAPS I household member with the lowest PID and all MAPS I household members living him/her.

*If (i) **all MAPS I household members** have moved out of the MAPS I dwelling, AND (ii) the head of household lives in the same EA or in its close vicinity (in Iganga and Mayuge) – either by himself or with the entire set or a subset of MAPS I household members:*

- Track the head living in the same EA or in its close vicinity. Once located, consider the household to be tracked.
- Mark the ID of the head and all MAPS I members living with the head.

*If (i) **all MAPS I household members** have moved out of the MAPS I dwelling, AND (ii) the head DOES NOT live in the same EA or in its very close vicinity (in Iganga or Mayuge), AND (iii) the spouse or other MAPS I household members live in the EA or in its close vicinity (in Iganga or Mayuge):*

Scenario #1:

- Attempt to track first the spouse living in the same EA or in its close vicinity. Once located, consider the household to be tracked.
- Mark the ID of the spouse and all MAPS I household members living with the spouse.

Scenario #2:

- If there was no spouse in MAPS I or the spouse is not in the same EA or in its close vicinity, locate the largest group of MAPS I household members living in the same EA or in its close vicinity. If there is a tie in group size, locate the group members that live closest to the MAPS I dwelling location. If there is yet another tie in terms of distance, locate the group that contains the lowest MAPS I PID. Once the group is located, consider the household to be tracked.
- Mark the ID of the MAPS I household member with the lowest PID and all MAPS I household members living him/her.

If all MAPS I household members moved out of the EA:

- If the household was a single-headed household in MAPS I...

Scenario #1:

- If the head is alive and living in Iganga or Mayuge, he/she is the tracking target. Once located, the household could be considered tracked.
- Mark the ID of the head and all MAPS I members living with the head.

Scenario #2:

- If the head is not alive or is not living in Iganga or Mayuge, and the household was NOT a one-member household in MAPS I, locate the largest group of MAPS I household members that are known to be living together in Iganga or Mayuge. If there is a tie in group size, locate the group members that live closest to the MAPS I dwelling location in Iganga or Mayuge. If there is yet another tie in terms of distance, locate the group that contains the lowest MAPS I PID. Once the group is located, consider the household to be tracked.
 - Mark the ID of the MAPS I household member with the lowest PID and all MAPS I household members living him/her.
- If the household had a head and a spouse in MAPS I...

Scenario #1:

- If the head and spouse are together elsewhere in Iganga or Mayuge, they are the tracking targets. Once located, consider the household to be tracked.
- Mark the ID of the head and all MAPS I household members living with the head and the spouse.

Scenario #2:

- If one of the partners is not alive, the remaining partner is the tracking target as

long as he/she is living in Iganga and Mayuge. Once located, consider the household to be tracked.

- Mark the ID of the head or the spouse and all MAPS I household members living with him/her.

Scenario #3:

- If one of the partners is not alive, the remaining partner is not living in Iganga and Mayuge, locate the largest group of MAPS I household members that are known to be living together in Iganga or Mayuge. If there is a tie in group size, locate the group members that live closest to the MAPS I dwelling location in Iganga or Mayuge. If there is yet another tie in terms of distance, locate the group that contains the lowest MAPS I PID. Once located, consider the household to be tracked.
- Mark the ID of the MAPS I household member with the lowest PID and all MAPS I household members living him/her.

Scenario #4:

- If the head and spouse have moved separately and are both known to be alive and living in Iganga and Mayuge, the head is the tracking target. Once located, consider the household to be tracked.
- Mark the ID of the head and all MAPS I household members living with him/her.

Scenario #5:

- If the head and spouse have moved separately and are both known to be alive but and only one of them is living in Iganga and Mayuge, the partner living in Iganga or Mayuge is the tracking target. Once located, consider the household to be tracked.
- Mark the ID of the head or the spouse and all MAPS I household members living with him/her.

Scenario #6:

- If the head and spouse have moved separately and only of them is alive and living in Iganga and Mayuge, the living partner living in Iganga or Mayuge is the tracking target.
- Once located, consider the household to be tracked.
- Mark the ID of the head or the spouse and all MAPS I household members living with him/her.

Scenario #7:

- If the head and spouse have moved separately and none of them is living in Iganga and Mayuge, locate the largest group of MAPS I household members that are living together in Iganga or Mayuge. If there is a tie in group size, locate the group members that live closest to the MAPS I dwelling location in Iganga or Mayuge. If there is yet another tie in terms of distance, locate the group that contains the lowest MAPS I PID. Once located, consider the household to be

tracked.

- Mark the ID of the MAPS I household member with the lowest PID and all MAPS I household members living him/her.

Scenario #8:

- If the head and spouse are both deceased, outside of Iganga and Mayuge or in an institution, locate the largest group of MAPS I household members that are living together in Iganga or Mayuge. If there is a tie in group size, locate the group members that live closest to the MAPS I dwelling location in Iganga or Mayuge. If there is yet another tie in terms of distance, locate the group that contains the lowest MAPS I PID. Once located, consider the household to be tracked.
- Mark the ID of the MAPS I household member with the lowest PID and all MAPS I household members living him/her.”

Appendix II - MAPS II Plot Selection Protocol

The following is an excerpt from the MAPS II field staff manual:

“Household is classified with pure stand cultivation status if they have at least one pure stand maize plot during the on-going agricultural season.

Similarly, a household is classified with intercropped cultivation status if they have at least one intercropped maize plot during the on-going agricultural season.

Note that a household may possess both pure stand and intercropped cultivation statuses.

If the household was of pure stand cultivation status in 2015, and the household is of pure stand cultivation status in 2016 in accordance with the definition above:

... If the parcel that was selected for crop cutting in 2015 is still in household's possession and has at least one maize plot that is of pure stand cultivation status, select a maize plot at random that is of pure stand cultivation status among the pure stand maize plots on the same parcel from 2015.

... If the parcel that was selected for crop cutting in 2015 either is not in household's possession OR is in household's possession BUT does not have any plots that is cultivated with maize (irrespective of cultivation pattern) OR is in household's possession and have at least one plot that is cultivated with maize BUT does not have at least one plot that is of pure stand cultivation status, select a pure stand plot at random among all the pure stand plots that are being cultivated by the household in 2016.

If the household was of intercropped cultivation status in 2015, and the household is of intercropped cultivation status in 2016 in accordance with the definition above:

... If the parcel that was selected for crop cutting in 2015 is still in household's possession and has at least one maize plot that is of intercropped cultivation status, select a maize plot at random that is of intercropped cultivation status among the intercropped maize plots on the same parcel from 2015.

... If the parcel that was selected for crop cutting in 2015 either is not in household's possession OR is in household's possession BUT does not have any plots that is cultivated with maize (irrespective of cultivation pattern) OR is in household's possession and have at least one plot that is cultivated with maize BUT does not have at least one plot that is of intercropped cultivation status, select a intercropped plot at random among all the intercropped plots that are being cultivated by the household in 2016.

If the household was of pure stand cultivation status in 2015, and the household is of not pure stand cultivation status and is of intercropped cultivation status in 2016 in accordance with the definition above:

... If the parcel that was selected for crop cutting in 2015 is still in household's possession and has at least one intercropped maize plot, select a maize plot at random that is of intercropped cultivation status among the intercropped maize plots on the same parcel from 2015.

... If the parcel that was selected for crop cutting in 2015 either is not in household's possession OR is in household's possession BUT does not have any plots that is cultivated with maize, select an intercropped plot at random among all the intercropped plots that are being cultivated by the household in 2016.

If the household was of intercropped cultivation status in 2015, and the household is of not intercropped cultivation status and is of pure stand cultivation status in 2016 in accordance with the definition above:

... If the parcel that was selected for crop cutting in 2015 is still in household's possession and has at least one pure stand maize plot, select a maize plot at random that is of pure stand cultivation status among the pure stand maize plots on the same parcel from 2015.

... If the parcel that was selected for crop cutting in 2015 either is not in household's possession OR is in household's possession BUT does not have any plots that is cultivated with maize, select a pure stand plot at random among all the pure stand plots that are being cultivated by the household in 2016.”

Appendix III – MAPS Sub-Plot and Full-Plot Crop Cutting Protocol

Sub-Plot Crop Cutting

The following is an excerpt from the MAPS I field staff manual. The MAPS II crop cutting protocol was identical to that of Round I, with one difference. In MAPS II, as opposed to setting up one 4x4m (divided into four 2x2m quadrants) and another 2x2m sub-plot on each plot, the enumerator set up only one 8x8m sub-plot (divided into four 4x4m quadrants).

"There are three aspects to this exercise – the first is conducted with the post-planting questionnaire and the last two are conducted at the time of harvest:

1. The first aspect is the selection of a random 4m x 4m crop-cutting subplot AND a random 2x2m crop-cutting subplot within the plot. Using rope, the 4mx4m subplot will be divided into four 2mx2m quadrants. The 4x4m subplot and the 2x2m subplot will be selected using random number tables. This will take place as part of the post-planting questionnaire.
2. The second aspect of this exercise is the harvesting of the maize once it is ready for harvest. This should be done at a time that is convenient for the farmer. It is very important that the farmer does not harvest the land before you arrive – therefore, please coordinate with the farmer and the local crop-cut monitor to learn the time at which he/she would like to harvest, and be sure to arrive without delay. The crop will be weighed at the time of harvest. Each of the five 2x2m quadrants will be weighed separately.
3. The third aspect is the drying of the crop. This will take place at a centralized location in Kampala. Dry weights of the maize harvest will be conducted by a separate team in Kampala.

The materials that you will need for use in this exercise are:

- Module K (Post-Planting Questionnaire) and Module M (Crop-Cutting Questionnaire)
- Compass
- Pre-measured 4m x 4m PVC pipe
- Pre-measured 2m x 2m PVC pipe
- Sticks (13) for Area Demarcation
- Measuring Tape
- Rope (32+ meters per household)
- Bags, Barcodes, Writing Materials
- Industrial Digital Weighing Scale

Module K and Module M: Module K will be completed as part of the post-planting visit, when the crop-cutting subplots are demarcated. Module M will be completed during the crop-cutting visit, when the subplots are harvested and the crop is weighed.

Compass: This is a device used for capturing geographic bearings in degrees (0⁰).

Pre-measured 4m x 4m PVC pipe: A set of PVC pipes that are pre-measured to create a 4x4 meter square will be provided to each enumerator to ensure the crop-cut area is precisely 4x4 meters.

Pre-measured 2m x 2m PVC pipe: A set of PVC pipes that are pre-measured to create a 2x2 meter square will also be provided to each enumerator to ensure the second crop-cut area is precisely 2x2 meters.

Sticks: These will be used to mark the four corners of the areas selected for crop cutting. Eight sticks will be used to mark the corners of the 4x4 meter subplot and four sticks to mark the corners of the 2x2 meter subplot.

Measuring Tape: This is a distance-measuring instrument marked in metric-units (segments), which will be used to determine the location of the areas in the plot.

Bags, Barcodes, Writing Materials: Each quadrant's harvest will be stored in bags that will be provided for sample transport. Each bag will be tagged with a water-resistant barcode sticker, whose duplicate will be placed inside the bag. You will be provided with pen and pencils for note taking.

Industrial Digital Weighing Scale: This will be used to weigh the harvested maize at the time of harvest (in grain form). Each of the five 2x2m quadrants must be weighed separately.

Procedure for Crop Cutting

We will be conducting crop cutting on a 4m x 4m subplot AND a 2m x 2m subplot of the maize plot. However, we will divide the 4mx4m area into four 2mx2m squares (also called quadrants). Therefore, there will be **a total of FIVE 2m x 2m quadrants**. The harvest of each quadrant will be recorded separately. Here, we describe in further detail each of the four main aspects to the crop cutting exercise.

You will first construct the 4m x 4m subplot by following steps 1 and 2 below. Only after demarcating the 4m x 4m subplot you will repeat steps 1 and 2 for the 2m x 2m subplot. **The 4x4 and 2x2 subplots may not overlap.**

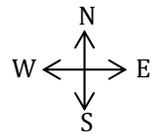
1) Crop Cutting Area Selection:

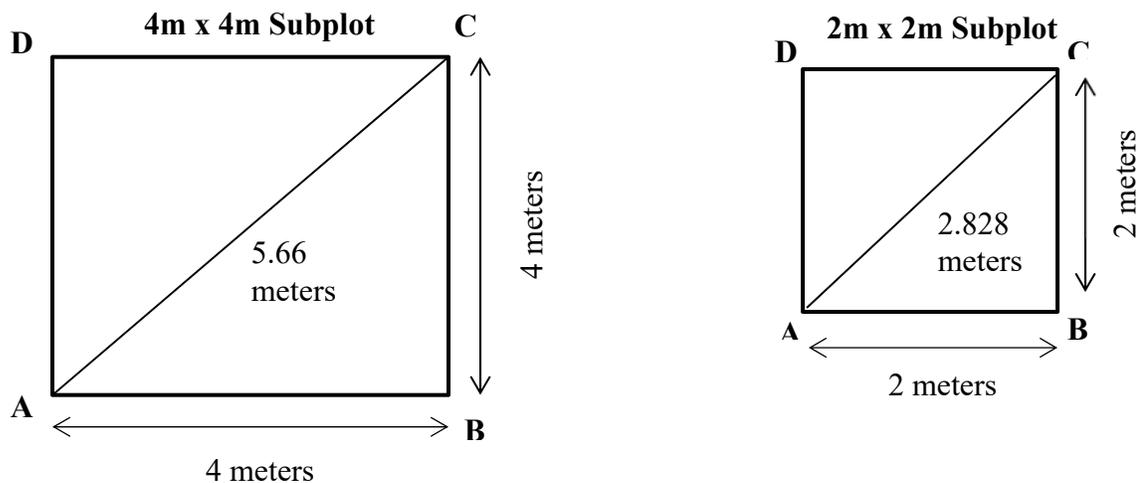
- a. Use Random Number Table #1 to identify the corner from which you will start. Use the first number in the random number table that matches one of the corners of the plot. The corner in which you started the area measurement, the northwest corner, is corner #1. Corner #2 is the next corner of the plot, moving around the plot clockwise.
- b. Measure the distance of the two sides along the selected corner with the measuring tape. Identify which is the longer side and which is the shorter side.
- c. Take the bearing from the start corner down the shorter side. Note this in your notebook.

- d. Use the Random Number Table #2 provided for this household. The first number should be the number of meters that you will walk along the length of the longer side of the plot. If the first number is larger than the length of the side, choose the next random number (and so on, until you find a number that is less than the length of the side). For example, if the length of the longer side is 25 meters and the first random number in the list is 28, move on to the next number.
- e. Beginning at your starting point and continuing along the longer side of the plot, walk the number of meters indicated by your random number.
- f. Turn into the plot so that your bearing is the same as the bearing you measured down the shorter side of the plot. This means you will be entering the plot parallel to the shorter side. Choose the next random number from Random Number Table #2 that is shorter than the length of the shorter side and walk the number of meters indicated by this second random number. You should be walking in a direction that is parallel to the shorter edge of the plot. Walk in a straight line. Try not to veer to the right or left to avoid shrubs or wet spots.
- g. The corner of the crop cutting subplot is located where your foot lands on the last step: this is point A.

2) Crop Cutting Subplot Demarcation:

- a. At point A, insert the first stick firmly into the ground, then turn your face to the east and lay the PVC pipe square in front of you (if you do not have PVC pipe with you, measure 4 meters *directly* to the east). One side of the PVC square should go from Point A to the east, which we call Point B. From Point B, the next corner of the PVC pipe should be to the north where we put Point C (if you do not have PVC pipe with you, measure 4 meters *directly* to the north).
- b. With the PVC square on the ground, insert sticks exactly at each corner.
- c. Tie a rope around all four sticks. Carefully dis-assemble and remove the PVC pipe square, leaving only the rope. The rope will stay on the subplot until the time of harvest.
- d. In order to make sure that the subplot size is correct, check to make sure that the diagonal line (Line A-C) is: (1) 5.66 meters on the 4m x 4m subplot and (2) 2.828 meters on the 2m x 2m subplot.



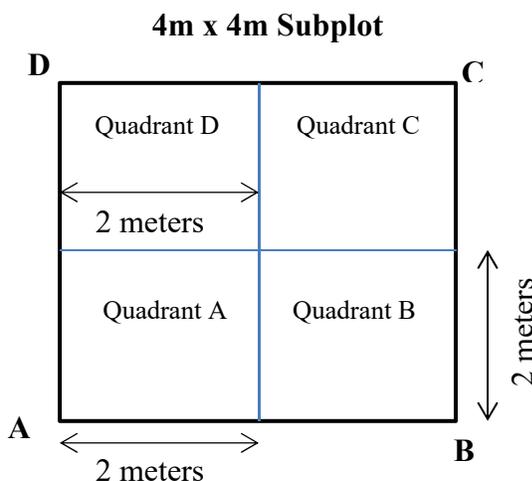


Note: If the random numbers obtained from the random table for long and short sides of the plot do not fall in the crop plot area, drop both random numbers and start over again. Each time when one or both of the random numbers fail to fall in the plot, drop both and start again until both random numbers fall on the plot. Also, if the 2x2 subplot overlaps with the 4x4 subplot, you must drop the random numbers and start again on the 2x2 subplot.

If there is an obstacle in one or more of the crop-cutting subplots, such as a large tree stump, a boulder, large ant hill, etc. **re-select** the subplot by starting with a new random corner.

If in one or more of the crop-cutting subplots there is maize damage, DO NOT re-select the subplot. Leave it as it and we will record the damage in the crop-cut questionnaire.

- e. **FOR THE 4x4 SUBPLOT ONLY:** The last step is to divide the 4m x 4m subplot into 4 equal quadrants. Measure 2m from each corner and enter a new stick. These new sticks will mark the middle of the rope on each side. Next, tie a piece of rope between the new sticks so that there are four (4) equal quadrants as in the example below. These four quadrants will be called quadrant A, B, C, and D. **When conducting the harvest, you will need to keep the crop from each quadrant separate.**



3) Harvest of Demarcated Section – Completed at the time of harvest

With the consent of the farmer, harvest all of the maize contained within the demarcated plot area, **keeping the crop from each quadrant separate**. Count the number of plants and cobs that are harvested from each quadrant. Once harvested and shelled, the grain should be weighed carefully using the digital scales, and the data recorded in the Module M of the Crop-Cutting Questionnaire. Remember, the crop from each quadrant **must** be weighed separately. The crop cut samples will be picked up on weekly supervision visits and delivered to the central drying location in Kampala. A separate team will then dry the crop for an additional time, weigh it again at a later date, and capture the moisture content at the time of the final weighing.

Full Plot Crop Cutting in MAPS II

Out of the initial target of 540 households, half of households in each of the pure stand and intercropped domains in each MAPS Round I EA were selected at random, prior to the start of the MAPS II fieldwork, to be subject to full-plot crop cutting, in addition to the 8x8m sub-plot crop cutting. This yielded a pre-full-plot crop cut sample of 282 households. Given the attrition dynamics in Round II and 31 households that were selected for full-plot crop cutting but that had to harvest their only plot prior to the crop cutting visit (with the exception of the crop cut sub-plot), the final sample that was subject to full plot crop cutting was composed of 214 households.

At the time of the harvest, the 8x8 crop cut sub-plot was harvested first, and both unshelled and shelled weights were taken, alongside shelled grain moisture readings for each of the 4x4m quadrant harvest. Subsequently, the rest of the plot was harvested for unshelled and shelled weight measurements and shelled grain moisture measurements, following the steps outlined below. For capturing unshelled and shelled weights tied to full-plot crop cut harvests, we used high-accuracy, digital, industrial HIWEIGH scales that were procured through and calibrated by the Uganda National Bureau of Standards. Each scale had a maximum load of 300 kilograms and a readability of 0.01 grams. We used DICKEY-john mini GAC moisture meters that were borrowed from the National Crop Resources Research Institute (NaCCRI). Each of the 3 survey teams had one scale and one moisture meter.

Due to different planting times, farmers were harvesting their grain at different times. Instead of concentrating on one EA at a time, the teams adopted a system of visiting an average of 2 to 3 EAs per day during the crop cutting fieldwork period, and conducting crop cuts on an average of 3 sampled plots across harvest-ready households. Each household was allocated, on average, four 100 kilogram bags to facilitate the full plot harvest. Every plot had 2 crop assistants recruited from the associated household to assist in the full-plot crop cut. Every crop assistant was paid 5000 Ugandan Shillings, approximately 1.5 USD. Further, each full-plot crop cut household received a tarpaulin (used for maize drying), in addition to a hoe and a panga knife that was provided to all MAPS II households.

The specific steps in the full-plot crop cuts were:

1. Visit the plot; and count and record the number of plants in each of the 4 quadrants of the crop cut sub-plot, and separately, in the rest of the plot.
2. Harvest, transport to the dwelling location, and count the cobs; and take unshelled weights, for the each of the 4 quadrants of the crop cut sub-plot, and separately for the rest of the plot, which would in turn be shelled by the farmer and household-specific crop cut assistants.

(The time commitment to Steps 1 and 2 for a 1-acre plot was approximately 2 hours. The crop cut assistants followed up with the households during the shelling period to ensure compliance with the instructions regarding (1) the separation of the harvests tied to the sub-plot quadrants versus the rest of the plot as well as other plots that may have been cultivated by the household, and (2) the prevention of household consumption of shelled maize until the weight and moisture measurements were taken.)

3. Visit the household in 2 to 3 days depending on the shelling progress; obtain shelled maize weights and moisture readings for the each of the 4 quadrants of the crop cut sub-plot, and for each 100 kg sack utilized by the household for storing the harvest tied to the rest of the plot.

Appendix IV – Inputs into MAPS Farmer-Reported Production Estimates

List of Allowable Farmer-Reported Production Measurement Units

Unit Code	Unit Description
1	Kilogram (kg)
2	Gram
4	Small cup with handle (Akendo)
9	Sack (120 kgs)
10	Sack (100 kgs)
11	Sack (80 kgs)
12	Sack (50 kgs)
13	Sack (unspecified)
20	Tin (20 lts)
21	Tin (5 lts)
22	Plastic Basin (15 lts)
29	Kimbo/Cowboy/Blueband Tin (2 kg)
30	Kimbo/Cowboy/Blueband Tin (1 kg)
31	Kimbo/Cowboy/Blueband Tin (0.5 kg)
37	Basket (20 kg)
38	Basket (10 kg)
39	Basket (5 kg)
40	Basket (2 kg)
63	Crate
64	Heap (Unspecified)
66	Bundle (Unspecified)
67	Bunch (Big)
68	Bunch (Medium)
69	Bunch (Small)
70	Cluster (Unspecified)
85	Number of Units (General)
99	Other Units (Specify)

List of Allowable Farmer-Reported Harvest Conditions

Condition Code	Condition Description
1	Green Harvested
2	Fresh/Raw Harvested
3	Dry At Harvest
4	Dry After Additional Drying

List of Allowable Farmer-Reported Harvest States

State Code	State Description
1	With Cob and Stalk
2	With Cob and Husk, Without Stalk
3	With Cob, Without Husk, Stalk
4	Grain
5	Other (Specify)

Appendix V – Construction of the Soil Quality Index and the Edge Effects

Soil Quality Index

The soil quality index was constructed following the guidance provided by Mukherjee and Lal (2014) in their comparison of three approaches to the computation of soil quality indices, including (i) a simple additive approach, (ii) a weighted additive approach, and (iii) a principal component approach. Indices following each of the three approaches were constructed, including numerous specifications of the principal component approach. Following bivariate analysis of each index and crop cutting yield estimates, the principal component approach was identified as the most appropriate method for this analysis.

Multiple indices were constructed using the principle component approach, with different combinations of soil properties. Ultimately, the index with the greatest predictive power with respect to crop cut yield was composed of organic carbon (%), soil electrical conductivity (an indicator of soil salinity), and pH. First, principal component analysis was estimated, followed by the identification of highly weighted soil properties in each component with an eigenvalue greater than or equal to 1. In accordance with the guidance set forth by Mukherjee and Lal (2014), the highest weighted soil property was maintained for each component (when greater than or equal to 1). In the case where the highest weighted soil property was within 10 percent of the weight of another property, both (or all) properties were maintained, provided the correlation of the properties was less than or equal to 0.60. Weights were then constructed for each of the retained variables, defined as the share of the variance explained by the respective component relative to the total cumulative explained variance. For detailed explanations, see Mukherjee and Lal (2014) and Andrews et al. (2002).

Linear scores for each retained soil property were determined by dividing all observations by the highest value in the sample for soil properties in which a higher value is more beneficial (organic carbon). Soil electrical conductivity and pH have an optimal range, whereby there are both lower and upper bounds. For these properties, the observations were split into those above and below the critical thresholds, as defined by Mukherjee and Lal (2014), with those below the threshold treated as though a higher value is preferred and those above the threshold treated as though a lower value is preferred. Linear scores are normalized over the sample and, therefore, range from zero to one. The final index is weighted mean of the three soil properties, constructed using weights identified in the principal analysis and the normalized soil property scores.

Edge Effects

The variables indicating the proximity of the crop-cutting sub-plot to the plot edge are constructed by overlaying the sub-plot border on the plot boundary. The sub-plot boundaries are re-constructed with the GIS tools based on the coordinate of the crop cut sub-plot starting corner. A variable is

then constructed for the share of the subplot that fell within a 4-meter internal buffer of the plot boundary. A separate variable is computed considering a 1-meter plot buffer for MAPS II, given the increase in the crop sub-plot area to 64 m² in this round.

In the case of MAPS I, since there are two crop cut sub-plots (a 4x4m sub-plot and a separate 2x2m sub-plot) and each may have a different share of the subplot within the buffer zone, an aggregated variable is necessary. This variable is computed as the sum of the shares of the sub-plots within the buffer zones, weighted by the total crop cut area of 20 m² that a given crop cut sub-plot accounted for:

$$\begin{aligned} & \textit{aggregate share in buffer zone} \\ &= (\textit{share of 4x4m subplot in buffer zone}) \times \frac{16m}{20m} \\ &+ (\textit{share of 2x2m subplot in buffer zone}) \times \frac{4m}{20m} \end{aligned}$$

The indicator for MAPS II is simply the share of the 8x8 meter subplot in the buffer zone.