Mission Impossible? Exploring the Promise of Multiple Imputation for Predicting Missing GPS-Based Land Area Measures in Household Surveys

T. Kilic; I. Yacoubou Djima; C. Carletto

The World Bank, Development Data Group, Italy

Corresponding author email: iyacouboudjima@worldbank.org

Abstract:

Research has provided robust evidence for the use of GPS as the new, scalable gold-standard in land area measurement in household surveys. Nonetheless, facing budget constraints, survey agencies often measure with GPS only plots within a given radius of dwelling locations. It is, subsequently, common for significant shares of plots not to be measured, and research has highlighted the selection biases resulting from using incomplete data. This study relies on nationally-representative, multi-topic household survey data from Malawi and Ethiopia with near-negligible missingness in GPS-based plot areas to validate the accuracy of a Multiple Imputation (MI) model for predicting missing GPS-based plot areas in household surveys. The analysis randomly creates missingness among plots beyond two operationally-relevant distance measures from the dwelling locations, conducts MI for each artificially-created dataset, and compares the distributions of the imputed plot-level outcomes, namely area and agricultural productivity, with the distributions of their true, observed counterparts. MI procedure results in imputed yields that are statistically undistinguishable from the true distributions with up to 82% and 56% missingness, respectively for Malawi and Ethiopia, for plots located more than 1 kilometer away from dwellings. The study highlights the promise of using MI for reliably predicting missing GPS-based plot areas.

Acknowledegment: The authors thank Tomoki Fujii and Alberto Zezza, Heather Moylan for their comments on the earlier versions of this paper.

JEL Codes: C83, Q15

#127

Mission Impossible? Exploring the Promise of MI for Predicting Missing GPS-Based Land Area Measures in Household Surveys

Abstract

Research has provided robust evidence for the use of GPS as the new, scalable gold-standard in land area measurement in household surveys. Nonetheless, facing budget constraints, survey agencies often measure with GPS only plots within a given radius of dwelling locations. It is, subsequently, common for significant shares of plots not to be measured, and research has highlighted the selection biases resulting from using incomplete data. This study relies on nationally-representative, multi-topic household survey data from Malawi and Ethiopia with near-negligible missingness in GPS-based plot areas to validate the accuracy of a Multiple Imputation (MI) model for predicting missing GPS-based plot areas in household surveys. The analysis randomly creates missingness among plots beyond two operationally-relevant distance measures from the dwelling locations, conducts MI for each artificially-created dataset, and compares the distributions of the *imputed* plot-level outcomes, namely area and agricultural productivity, with the distributions of their *true, observed* counterparts. MI procedure results in imputed yields that are statistically undistinguishable from the true distributions with up to 82% and 56% missingness, respectively for Malawi and Ethiopia, for plots located more than 1 kilometer away from dwellings. The study highlights the promise of using MI for reliably predicting missing GPS-based plot areas.

JEL Codes: C53, C83, Q12, Q15.
Keywords: Survey Methodology, Global Positioning System (GPS), Land Area Measurement, Missing Data, Multiple Imputation, Malawi, Sub-Saharan Africa.

1 Introduction

Land area is a fundamental input into statistical and economic analyses linked to agriculture, inequality and land registration, titling and redistribution programs. The Sustainable Development Goal (SDG) Targets 2.3 and 2.4 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, respectively. Both targets are associated with indicators[1] that rely on land area information sourced from household or farm surveys, and research has demonstrated the importance of accurate land area measurement for accurate measurement and analysis of land productivity (Carletto, et al., 2013) (Carletto, et al., 2015).

While data collection on smallholder production systems has traditionally relied on self-reported land areas, this is problematic, particularly in the African context, which is characterized by the high incidence of smallholder farming and the fragmentation of farms into multiple parcels with irregular shapes and without formal titles. Several reasons may contribute to the inaccuracy in self-reported land areas. First, farmers may knowingly overstate or understate their landholdings for strategic reasons that may relate to access to development programs and/or taxation. Second, there is a natural tendency to round off numbers and provide approximations, which leads to heaping of the data around discrete values. Third, geography, particularly slope, can influence the way farmers assess distance and area. Fourth, the use of non-standard measurement units and within-country variation in the type and standard unit equivalence of these units complicate the compilation of conversion factors for land area measurement. Fifth, the magnitude and direction of the measurement error in self-reported land areas have been shown repeatedly to be systematically associated with observable plot, household and respondent attributes.

These reasons, combined with (i) the validated accuracy of GPS-based land area measurement in household survey experiments in Ethiopia, Nigeria, and Tanzania (Zanzibar) (Carletto, et al., 2016), and (ii) the ever-increasing affordability and accuracy of handheld GPS devices makes GPS-based land area measurement a desirable alternative for household and farm surveys in countries dominated by smallholder agricultural production. However, with the emergence of GPS-based area measurement as the new, scalable gold-standard for household and farm surveys, a key drawback is related to the operationalization of the technology. To reduce transportation costs, keep household interview durations within reasonable limits, and avoid the difficulty of asking respondents to accompany enumerators to agricultural plots that are situated far from dwelling locations, survey implementing agencies often require enumerators to only obtain GPS-based area measures for plots within a given radius of dwelling locations. Thus, non-ignorable shares of area measures are missing in public use datasets. For instance, among the selected national, multi-topic panel household surveys that are supported by the World Bank Living Standards Measurement Study – Integrated Surveys on Agriculture (LSMS-ISA) program, the rate of missingness in GPS-based plot areas range from 13 (Nigeria) to 44 percent (Uganda), as shown in Table 1.

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

The missing data, in turn, may limit the operational relevance and the analytical value of GPS-based area measurement, given the potential biases introduced by missingness.

Recognizing the need to address the problem of missing data for increasing the usability of household survey data, (Kilic, et al., 2017) use LSMS-ISA data from Tanzania and Uganda to show that the missing GPS-based plot areas constitute a non-random subset of the unit-record data, but that the missing data can be simulated by Multiple Imputation (MI). In their analysis of plot-level agricultural productivity, the authors document the non-trivial effects of using the completed datasets following MI.

Given potential of MI for imputing the missing GPS-based land areas and the importance of rigorously addressing missingness for productivity estimation, this paper takes on the challenging task of determining thresholds for acceptable rates of item non-response in plot areas in large-scale surveys that adopt GPS technology for land area measurement. To do so, we work with national household survey data from Malawi and Ethiopia that exhibit near-negligible rates of missingness in GPS-based plot areas, and use these datasets as validation samples to gauge the accuracy of an MI application to predict missing GPS-based land areas.

The use of actual data collected as part of large-scale household surveys that had adopted GPS-based area measurement is key to the operational relevance of our research. We set up our empirical framework as to identify the limits to simulation accuracy and provide recommendations for capturing the minimum set of required data for robust statistical analyses relying on plot areas. Specifically, in both datasets, we

i. create artificial missingness in GPS-based plot areas at random at a rate of 1 to 100 percent, at an increment of 1 percent, among plots that are above two distance thresholds, namely a distance of greater than 500 meters or 1 kilometer from the dwelling unit,
ii. construct an imputation model for missing GPS-based plot areas following (Kilic, et al., 2017),
iii. conduct MI based on each unique data set under a specific simulated degree of missing observations beyond the two different distance thresholds,
iv. compare the distributions of two outcomes, namely plot area and plot-level agricultural productivity, based on the *imputed* GPS-based plot areas with the distributions of the same variables based on the *observed* area measures for the same plots, and lastly
v. identify the missingness threshold beyond which MI yields at least 1 imputed distribution out of a total of 50 imputations that is statistically different from the observed distribution at the 5 percent level.

The headline finding is that in Malawi, MI can produce imputed yields that are statistically undistinguishable from the true distributions with up to 82 percent missingness in plot areas that are further than 1 kilometer with respect to the dwelling location. The comparable figure in Ethiopia is 56 percent. These rates correspond to overall rates of missingness of 23 percent in Malawi and 13 percent in Ethiopia. If one sets the distance threshold at 500 meters, the imputed yields are statistically undistinguishable from the true distributions with distant plot missingness up to 45 percent in Malawi and

36 percent in Ethiopia, translating, respectively, into overall tolerable missingness rates of 21 percent and 15 percent.

The paper is organized as follows. Section 2 describes the data. Section 3 presents the empirical approach. Section 4 discussed the results. Section 5 concludes.

2 Data

The Malawi Third Integrated Household Survey 2010/2011 (IHS3), and the Ethiopia Socioeconomic Survey Wave II 2013/2014 (ESS2), which were conducted respectively by the Malawi National Statistical Office (NSO) and the Central Statistics Agency (CSA) of Ethiopia inform our analysis. Both surveys were implemented under the Living Standards Measurement Study – Integrated Surveys on Agriculture (LSMS-ISA) program.

The IHS3 data were collected within a two-stage cluster sampling design, and are representative at the national, urban/rural, regional, and district levels, covering 12,271 households in 768 enumeration areas (EAs). ESS2 is part of a long-term project to collect panel data. It covered all regional states including the capital, Addis Ababa. Much of the sample is comprised of rural areas as it was carried over from ESS1. The survey is representative at the national, urban/rural and, 6 strata (4 regions plus Addis Ababa and the other regions) covering 5,262 households in 433 EAs.

In terms of questionnaire instruments, the IHS3 and the ESS2 both had Household, Agriculture, and Community Questionnaires. In each setting, the sample households were administered a multi-topic Household Questionnaire that collected individual-disaggregated information on demographics, education, health, wage employment, nonfarm enterprises, anthropometrics, and control of income from non-farm income sources, as well as data on housing, food consumption, food and non-food expenditures, food security, and durable and agricultural asset ownership, among other topics. In addition, agricultural households received the Agriculture Questionnaire, which solicited information on land areas, manager/holder identification, physical characteristics, labor and non-labor input use, and crop cultivation and production at the plot-level.

Further, it is important to note that the IHS3 and the ESS2 make a clear distinction between a *parcel* and a *plot*. 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.[2] Further, agricultural production data were collected for the two main agricultural seasons in each survey. Handheld global positioning system (GPS)-based locations and land areas of the plots were recorded, permitting us to link household- and plot-level data to outside geographic information system (GIS) databases.

The IHS3 required GPS-based area measurement of all plots that are owned and/or cultivated by the sampled households, within 2 hours of travel with respect to the household location, regardless of mode of transportation. For the distant plots, the field teams were advised to cluster them in accordance with their location, and to visit them in a coordinated fashion by using the team vehicle. For the sub-sample of IHS3 households that were visited twice, the first visit data were also reviewed, and the missing GPS-based plot areas were fed forward to the second visit interviews for potential capture by the field teams. While the first visit constraints leading to missing data still applied to most of these households during the second visit, the continuing emphasis on increasing the volume of GPS-based plot area measures did result in additional data capture. On the other hand, the ESS2 instructed the enumerators to take GPS-based area measures of all plots that are owned and/or cultivated by the sampled households, irrespective of distance. For plots less than 40 square meters, the enumerators measured areas by traversing, instead of GPS units. The overall rates of missingness in GPS-based plot areas were considerably low in both settings: 3.8 percent in Malawi and 6.2 percent in Ethiopia. These are in fact the lowest levels observed among the surveys supported by the LSMS-ISA program.

Our analysis assumes both data sets to be complete and representative of the true distributions of interest, and is subsequently conducted using plots for which GPS based-land area measurements are available.[3] Table 2 shows the distribution of plots according to their distance from the dwelling for both datasets. Table 3 presents the summary statistics based on the IHS3 and the ESS2, including the plot-level means for the entire sample; for the sample within 1 kilometer of the dwelling; and for the sample that lie outside of the 1 kilometer radius of the dwelling. Table 4 accomplishes the same objective but for the

[2] Parcel-level GPS-based area estimation could serve other objectives, such as surveying of land for land registration or titling programs or for land ownership measurement. An open empirical question is whether the extent to which parcel-area measurement could be reliably backed from aggregation of within-parcel, plot area measures – an exercise that will be mediated by the precision with which parcel and plot boundaries are established in the field prior to GPS-based area measurement.

[3]We cannot work with approximately 50 percent of the ESS2 plots in the public use data since the CSA ancillary dataset with the conversion factors for the non-standard land area measurement units (to express farmer-reported plot areas in hectares) does not include conversion factors for all non-standard measurement units. This limitation further underscores the importance using GPS-based land area measurements. Going forward, the ESS2 can be used to update the referenced ancillary dataset of conversion factors. Prior to the validation exercise, we elected not to update the ancillary dataset using the ESS2 since the imputation model performance would have improved dramatically in a mechanical manner. Further, the overwhelming majority of the predictors that we use in the validation exercise based on the ESS2 data do present statistically significant differences across the plots depending on whether land area conversion factor is available. These predictors are included in the imputation model, and to the extent that they are correlated with observed and unobserved attributes that predict the likelihood of a farmer-reported plot area with a missing conversion factors, the ESS2 sample that we end up focusing on should be deemed satisfactory for validation purposes.

samples split by the alternative, 500-meter, distance threshold. We provide the differences between the sample means, and note when a given mean difference is statistically significant.

Several noteworthy findings emerge from Tables 2, 3 and 4. First, the distribution of plots per distance threshold is quite similar across the two countries. Between 54 and 60 percent of the plots are within 500 meters and between 72 to 77 percent are within 1.0 km. Second, the plots within the distance threshold tend to be of significantly smaller areas than the plots beyond that threshold. Third, several important plot and household level characteristics which are expected to be associated with productivity related outcomes, display statistically significant differences by distance threshold status. As also noted by (Kilic, et al., 2017), these observations highlight the importance of systematically addressing missingness in GPS-based plot areas, if such GPS data are to be used in a robust fashion.

3 Empirical Approach

3.1 Artificial Missingness Creation

The first step in our analysis is to generate missing GPS-based plot areas in a way that would be similar to real-life field experience. Missing GPS-based plot areas measurements are often tied to numerous field logistics and cost constraints. However, the variable that underlies the lion share of missing GPS-based plot areas in household survey operations is the plot distance from the dwelling or the location with respect to the EA boundaries. As noted above, the IHS3 instructed the enumerators to measure all plots within 2 hours travel time from the dwelling locations, while the ESS2 required the measurement of all plots, with the exception of those less than 40 square meters, irrespective of distance/travel time. For a more time and/or budget constrained operation, a lower threshold for GPS based land areas measurements could be enforced.

Our study uses 500 meters and 1 kilometer as the distance thresholds beyond which 1 to 100 percent of GPS-based plot area observations are artificially and randomly tagged to be missing in an increment of 1 percentage point. The distance variable underlying the thresholds is the Euclidean (crow-fly) distance between the geo-referenced plot and dwelling location.[4] To get sense of the time requirements associated with visiting plot locations that are below versus above the chosen distance thresholds, consider, for instance, the walking time associated with the inclination-adjusted minimum cost distance between dwelling and plot locations in Malawi. For plots that are within the 500 meter and within the 1 kilometer threshold, the average walking time is 4 minutes and 6 minutes, respectively. Conversely, for plots that

[4] Other geospatial measures of the plot distance to the dwelling were considered, including the estimated minimum cost distance that considers topography; the walking time associated with the minimum cost distance; and the inclination-adjusted measures of these two variables. The weighted pairwise correlation between any of the alternatives and our Euclidean distance measure is above 99 percent, and our results are robust to the use of these alternative distance measures.

are outside the 500 meter and outside the 1 kilometer threshold, the average walking time is 33 minutes and 47 minutes, respectively.

3.2 Multiple Imputation

The second step of our approach is to use Multiple Imputation (MI) to fill the gaps that we artificially create in the GPS-based plot area measures. MI, first proposed by (Rubin, 1987), is a Monte Carlo technique that replaces missing values for a given variable with $m > 1$ simulated alternatives. MI typically consists of three steps: (i) *m* imputations (i.e. *m* complete datasets) are generated based on an *imputation model* that encompasses a vector of observable covariates that predict the missingness in a given variable, (ii) statistical analysis is performed separately with each of the *m* complete datasets, and (iii) the results obtained from *m* complete data analyses are combined into a single set of multiply-imputed parameter estimates and standard errors.

The conditions under which valid inferences could be obtained from missing data is laid out in Rubin's (1987) seminal work on MI. The procedure assumes that data are missing at random (MAR) and that missing data could be predicted based on observable attributes underlying missingness. While the MAR assumption is not empirically testable, the limits of its tenability could be assessed in our study.

In building the imputation model, the literature (Rubin, 1996) or (van Buuren, et al., 1999) advises to include as explanatory variables: (i) the variables appearing in the analysis model that features the multiply-imputed variable(s), (ii) the variables that are known to have influenced the occurrence of missing data, and other variables for which the distributions differ between the response and non-response groups, (iii) the variables that explain a considerable amount of variance of the multiply-imputed variable(s) and that help to reduce the uncertainty of the imputations, and (iv) the variables with information on the features of the complex survey design, including stratum and cluster identifiers, and sampling weights.

In their MI application to missingness in GPS-based land areas in Tanzania and Uganda, (Kilic, et al., 2017) attempt to provide support for the MAR assumption by (i) detailing the field work processes underlying the missing data, (ii) providing insights from their field experience and interactions with the survey teams, (iii) systematically documenting the established guidelines on imputation model specification, and (iv) including in the imputation model explanatory variables that influence the occurrence of missing data; that have different distributions between the response and non-response groups; that explain a considerable amount of variance of the multiply-imputed variable; and that include information on the survey design. Our approach to specifying the imputation model mirrors that of (Kilic, et al., 2017). A key covariate that is included in the imputation model and that is both a powerful predictor and an alternative measure of the GPS-based plot area is the farmer-reported plot area. The availability of this variable distinguishes our study as well as (Kilic, et al., 2017) from other studies that have employed MI to tackle item non-response.

For illustration, Table 5 and Table 6 show the details of the Ordinary Least Squares (OLS) imputation model for Malawi and Ethiopia, respectively. In addition to farmer-reported plot area, we include plot manager, household and other plots attributes as predictors. The model specification differs slightly between the IHS3 and the ESS2 depending of the availability of the variables or the specificity of the data set. For example, the raw data on farmer-reported plot areas could have been expressed in non-standard measurement units in the ESS2, as such we add dummy variables for these units in the imputation model for Ethiopia.

We estimate the imputation model using each dataset that is created by a given distance threshold-artificial missingness combination. While the results confirm that the predictions are essentially driven by the farmer-reported plot area, the more comprehensive model improves the accuracy and precision of our predictions. As pointed out by (Kilic, et al., 2017), it is worth emphasizing that the imputation model neither intends to provide a parsimonious description of the data nor attempts to portray structural relationships among variables. Instead, it attempts to be as comprehensive as possible to minimize any bias that could stem from omitting variables that might be relevant to the pattern of missingness or the subsequent analysis. "The possible lost precision when including unimportant predictors is usually viewed as a relatively small price to pay for the general validity of analyses of the resultant multiply-imputed database" (Rubin, 1996).

In multiply imputing missing values that have been artificially created in each scenario, we fit plot-level OLS regression models with the GPS-based plot area as the dependent variable and obtain linear predictions for all plots in the dataset. Under the partially parametric method of predictive mean matching (PMM), we use the linear prediction as a distance measure to form a set of 5 nearest neighbors chosen from the plot sample with GPS-based area measures, and randomly pick one of the neighbors whose observed GPS-based plot area value replaces the missing value for the incomplete case at hand.[5]

The imputation is carried out 50 times[6] to reduce the potential sampling error due to imputation, and 50 complete datasets are generated. The posterior estimates of the model parameters are obtained using sampling with replacement, which is standard practice when the asymptotic normality of parameter estimates is suspect.[7] By drawing from the observed data, PMM preserves the distribution of observed values in the missing part of the data, which makes it more robust than the fully parametric regression approach. In total, we generate 50 complete datasets of GPS-based land plot areas for each of rate missingness (100) for each distance threshold for each country datasets. These data sets are used to assess the tolerable rates of missingness, as explained below.

[5] The results are robust to using linear regression, as opposed to PMM. The number of nearest neighbors in the PMM framework is inversely related to the correlation among imputations. While high correlation may increase the variability in MI point estimates, low correlation may increase the bias in MI point estimates. The literature does not provide definitive guidance on the decision regarding the number of nearest neighbors, but the results are robust to the specification of ten nearest neighbors, with or without bootstrapping.

[6] The results are robust to performing 100 imputations instead.

[7] The results are robust to sampling estimates from the posterior distribution of model parameters, as opposed to bootstrapping.

3.3 Assessing the tolerable rates of missingness in GPS-based plot areas

In order to assess the performance of the imputation model, we compare, the distributions of the *true*, *observed* versions of key variables that rely on GPS-based plot areas with the distributions of their completed (observed plus imputed) counterparts. The key outcomes that our assessment focuses on is GPS-based plot area and plot-level agricultural productivity, which is measured as the quantity or value of crop harvested based on farmer-reporting (the numerator) over cultivated land (the denominator). As discussed earlier, plot-level agricultural productivity is of policy relevance.

Given the nature of the problems to which MI is applied, it appears difficult for analysts to verify the appropriateness of their imputation procedures. Imputation values are guesses of unobserved, unknown values (Abayomi, et al., 2008). In this study, however, missingness is artificially created such that the true values are known. This allows direct comparison of the distributions of the observed vs. the completed data. Numerically, the comparison of the empirical distributions is done using the Kolmogorov-Smirnov (KS) test for each outcome variable for the different level of missingness, raising the flag when there's statistically significant differences at the 5 percent level[8] for at least 1 of the 50 imputations generated. As noted by (Abayomi, et al., 2008), there is no reason to suppose that setting a 5 percent level of significance will be appropriate when producing a MI diagnostic through density comparisons. However, it is useful to start with this rule and further examine the results.

4 Results

The results of our simulations are illustrated in Figure 1. Each panel shows the results for one threshold of one dataset. The first panel, for example, shows the result for the IHS3 when we impose a threshold of 1 kilometer. In each panel, the y-axis shows the number of imputations out of 50 for which, the KS test indicates that the distribution of the relevant outcome variable derived from the imputed GPS-based land area, was statistically indistinguishable from its observed counterpart. We also highlight the tolerable rates of missingness with a vertical line. The x-axis, on the other hand, shows the percentage of simulated missing GPS-based plot areas measurements beyond a given distance threshold. Three general observations emerge from Figure 1.

First, for low rates of missingness, all 50 imputations are statistically indistinguishable from the true distribution. As the rate of missingness increases, this count starts to decrease until only a small number (between 0 and 10) of the imputations appear to have distributions that are not statistically different from the observed true distribution. Second, within each data set, the tolerable rate of missingness is lower for

[8] The *p*-values for the test are approximate. The imputations are generated from the observed data. Hence, the empirical distributions are not independent of the observed data.

500 meters than it is for 1 km. Third, plot-level agricultural productivity is more sensitive to missingness than plot area (i.e. the tolerable rate of missingness is reached earlier in the case of the latter).

The first and second observations confirm the expectations anchored in the descriptive analyses discussed in Section 2. Plots that are further from the dwelling are inherently different from the ones that are closer. Thus, as missingness increases, the pool of plots with similar characteristics (and thus comparable areas) to choose from gets smaller, and it is understandable that the distribution differs substantially. The third observation is also foreseen: land area being the denominator of the formula for yield, a small deviation of the imputed values from the observed land values brings about a relatively more important deviation in the yield estimates obtained from them. Consequently, the yields calculated from the imputed land areas differ substantially from the true yields at lower rates of missingness.

We now compare the results obtained in the different panels depicted in Figure 1. For convenience, the tolerable rates of missing GPS-based plot areas are summarized in Table 7. Along with the tolerable rates in terms of the percentages of plot areas observations that could go missing beyond a given distance threshold, we report the corresponding overall rates of missingness in parentheses. In the discussion that follows, we focus on the discussion of the results pertaining to plot-level agricultural productivity, given the policy relevance of the outcome and its lower tolerance to missingness vis-à-vis plot area.
The results obtained with the 1 kilometer threshold are very encouraging. In IHS3, the MI procedure can produce imputed yields that are statistically undistinguishable from the true distributions at rates of up to 82 percent. For Ethiopia, the comparable figure is 56 percent, indicating that the plot-level agricultural productivity estimation is more sensitive, compared to Malawi, to missingness among the GPS-based plot areas that are beyond the 1 kilometer threshold. These rates translate into overall tolerable missingness rates of 23 percent and 13 percent in Malawi and Ethiopia, respectively.

As noted above, we get lower tolerable rates of missingness among distant GPS-based plot areas when we lower the threshold from 1 kilometer to 500 meters. In this case, the MI procedure can produce imputed yields that are statistically undistinguishable from the true distributions with up to 45 percent missingness among distant GPS-based plot areas in Malawi. The comparable figure is 36 percent for Ethiopia. These rates translate into overall tolerable missingness rates of 21 percent and 15 percent in Malawi and Ethiopia, respectively.

The cross-country differences in tolerable missingness rates are likely in part tied to the differences in farm organization.[9] On the one hand, the average plot size in hectares in Malawi (0.4) is twice as much as the comparable statistic in Ethiopia (0.2), as reported in Table 3. On the other hand, the household-level average number of plots per holding in Ethiopia (11.7) is more than six times the comparable figure in Malawi (1.9). While the spatial distribution of the plot samples across the distance intervals in Table 2 are comparable across the two settings, the average plot distance from the dwelling is 2.19 kilometers in Malawi, with a 95 percent confidence interval of 1.91-2.47, versus 1.10 kilometers in Ethiopia with a 95

[9] Unless otherwise stated, the statistics in this paragraph are not reported in any of the tables, but have been computed based on the same datasets used for analysis.

percent confidence interval of 0.76-1.43. The plot distance from the dwelling further exhibits cross-country distributional differences that are statistically significant at the 1 percent level.
Finally, Table 8 presents country-specific multiply-imputed mean versus true mean comparisons for plot-level area and agricultural productivity, following MI at identified tolerable rates of missingness above the distance thresholds as reported in Table 7. Irrespective of the distance threshold and country in question, the root mean square error for plot area is close to zero and the difference between the MI mean and the true mean as a percentage of the true mean does not exceed 1.5 percent. For plot-level agricultural productivity, we have more promising findings in Malawi compared to Ethiopia. In Malawi, for instance, at 82 percent missingness above the 1 kilometer threshold, the difference between the MI mean and the true mean as a percentage of the true mean stands at 7.5 percent. The comparable statistic for Ethiopia is 40.4 percent. These findings underscore the relative sensitivity to missingness of plot-level agricultural productivity measures vis-à-vis plot area, and the fact that this sensitivity is likely to vary by country and production system complexity, as in this study.

5 Conclusion

This paper provides further evidence that combining GPS-based plot areas measurements with farmer-reported plots areas in a sound Multiple Imputation (MI) application can result in reliable simulations of missing GPS-based plot areas. While the idea was first pursued by (Kilic, et al., 2017) using data from Tanzania and Uganda, our analysis extends the pursuit with data from Malawi and Ethiopia featuring negligible levels of missing GPS measurements. By artificially simulating the missingness in otherwise assumed-to-be-complete data from these two settings, we compare the MI-based predictions to the true, observed values and gauge the levels of missingness in GPS-based land area measurements that can be handled with MI without compromising the robustness of key land area related statistics.

Among the outcome variables of interest, plot-level agricultural productivity, as measured by maize yield in Malawi and total harvest value per land area in Ethiopia, is more sensitive to missingness. Still, in Malawi, MI can produce imputed yields that are statistically undistinguishable from the true distributions with up to 82 percent missingness in plot areas that are further than 1 kilometer with respect to the dwelling location. The comparable figure in Ethiopia is 56 percent. In other words, if only 18 percent of the distant plot areas in Malawi and 44 percent of the distant plots in Ethiopia were randomly selected for GPS-based area measurement, one can generate reliable, imputed plot-level measures for area and agricultural productivity. If implemented in future surveys, this would clearly result in significant savings in terms of time and resources.

However, since the tolerable missingness rates vary by country, distance threshold and outcome variable, prior to scaling up, it is imperative to replicate similar analyses using other survey data that exhibit low rates of missingness in GPS-based plot areas in order to converge on comprehensive operational guidelines for survey practitioners. Nevertheless, the potential of using MI for complementing missing GPS measurements is evident and should be pursued whenever possible.

Finally, although dealing with missingness empirically in the post-fieldwork period is usually an option, there is no substitute for good fieldwork to prevent unwarranted missing measurements as much as possible. Thus, we would advise countries to follow a combination of (i) well-supervised field practices aimed at reducing missingness, as exemplified in Section 2, and (ii) sound MI applications to fill the data gaps that will still be unavoidable to a degree.

Table 1: Rates of Missingness in GPS-Based Plot Areas in Selected Datasets Generated by the World Bank Living Standards Measurement Study – Integrated Surveys on Agriculture (LSMS ISA) & Survey Instructions on the Required Spatial Coverage of GPS-Based Plot Area Measurements

Survey	***Rate of Missingness***	***Required Spatial Coverage of GPS-Based Plot Area Measurements***
Niger Enquete Nationale sur les Conditions de Vie des Menages et l'Agriculture 2011	29%	Measure all plots in the same enumeration area as the household.
Nigeria General Household Survey - Panel 2012/2013	13%	Measure all plots in the same district of the household and within 3 hours of travel, regardless of mode of transportation.
Tanzania National Panel Survey 2010/2011	22%	Measure all plots within 1 hour of travel from the household, regardless of mode of transportation.
Uganda National Panel Survey 2011/2012	44%	Measure all plots in the same enumeration area as the household.

Table 2: Plot Distribution Based on the Euclidean Distance from Household

	Malawi (IHS3)			**Ethiopia (ESS2)**		
Distance Interval	**Frequency**	**Percentage**	**Cumulative Percentage**	**Frequency**	**Percentage**	**Cumulative Percentage**
[0.0, 0.5 Km)	9,798	53.67	53.67	12,282	61.51	61.51
[0.5, 1.0 Km)	3,363	18.42	72.09	3,070	15.38	76.89
[1.0, 2.0 Km)	2,888	15.82	87.91	2,455	12.30	89.19
[2.0, 3.0 Km)	755	4.14	92.05	862	4.32	93.50
[3.0, 5.0 Km)	404	2.21	94.26	537	2.69	96.19
[5.0, 10.0 Km)	306	1.68	95.94	342	1.71	97.91
[10.0, ~ Km)	742	4.06	100.00	418	2.09	100.00
Total	18,256	100.00		19,966	100.00	

Table 3: Selected Plot-Level Means by Plot Distance to Household (Above versus Below 1 Kilometer)

	Malawi (IHS3)				Ethiopia (ESS2)			
	Entire sample	**Sample [d < 1km]**	**Sample [d > = 1km]**	**x[d <1km] - x[d >= 1km]**	**Entire sample**	**Sample [d < 1km]**	**Sample [d > = 1km]**	**x[d <1km] - x[d >= 1km]**
Observations (Plots)	18,256	13,161	5,095		19,966	15,352	4,614	
Plot Areas								
GPS-based plot area (Ha)	0.394	0.383	0.420	-0.037***	0.197	0.177	0.261	-0.084***
Farmer-reported plot area (Ha)	0.414	0.403	0.440	-0.036***	0.193	0.175	0.251	-0.075***
Yields								
Maize yield (Kg/Ha)	1,693	1,694	1,692	2				
Value of output/Ha					29,303	31,575	22,447	9,128
Plot Manager Characteristics								
Female †	0.261	0.267	0.246	0.021	0.153	0.158	0.137	0.021
Age (Years)	43.147	43.511	42.273	1.238***	46.817	47.220	45.472	1.748*
Education (Years)	5.028	4.934	5.252	-0.318***	1.874	1.952	1.614	0.338
Household Characteristics								
Household size	4.934	4.871	5.086	-0.215***	6.476	6.491	6.427	0.064
# of HH members - [0,5]	0.981	0.974	0.998	-0.024	0.916	0.920	0.904	0.016
# of HH members - [6,14]	1.396	1.369	1.461	-0.093**	1.932	1.930	1.940	-0.010
# of female HH members - [15,39]	0.901	0.879	0.953	-0.074***	1.111	1.102	1.143	-0.041
# of male HH members - [15,39]	0.837	0.819	0.881	-0.062**	1.212	1.210	1.217	-0.007
# of female HH members - [40,59]	0.270	0.270	0.268	0.002	0.386	0.394	0.363	0.031
# of male HH members - [40,59]	0.269	0.262	0.287	-0.024	0.391	0.387	0.404	-0.017
# of HH members – 60 & above	0.280	0.297	0.238	0.059***	0.527	0.548	0.456	0.092*
Household consumption expenditures per capita	50,431	48,494	55,087	-6,593***	5,723	5,804	5,453	350
Number of plots in the holding	2.374	2.359	2.410	-0.051	15.857	16.086	15.091	0.995
Plot Characteristics								
Owned by household †	0.904	0.917	0.872	0.045***	0.866	0.888	0.794	0.094***
Use of hired labor †	0.223	0.195	0.290	-0.095***	0.057	0.050	0.082	-0.032***
Use of organic fertilizer †	0.116	0.122	0.101	0.021***	0.183	0.213	0.085	0.128***
Use of inorganic fertilizer †	0.618	0.623	0.607	0.016	0.404	0.415	0.369	0.047
Irrigated †	0.005	0.005	0.006	-0.001	0.016	0.017	0.011	0.006
Soil quality good †	0.467	0.453	0.503	-0.050***	0.327	0.329	0.319	0.010
Soil quality poor †	0.113	0.112	0.116	-0.004	0.173	0.168	0.188	-0.019

Note: † denotes a dummy variable. *** p<0.01, ** p<0.05, * p<0.1. Sample of plots within a 1 kilometer radius is the comparison group for the tests of mean differences.

Table 4: Selected Plot-Level Means by Plot Distance to Household (Above versus Below 500 meters)

	Malawi (IHS3)				Ethiopia (ESS2)			
	Entire sample	**Sample [d < 500m]**	**Sample [d >= 500 m]**	**x[d <500m] - x[d >= 500m]**	**Entire sample**	**Sample [d < 500m]**	**Sample [d >= 500 m]**	**x[d <500m] - x[d >= 500m]**
Observations (Plots)	18,256	9,798	8,458		19,966	12,282	7,684	
Plot Areas								
GPS-based plot area (Ha)	0.394	0.377	0.412	-0.035***	0.197	0.163	0.249	-0.086***
Farmer-reported plot area (Ha)	0.414	0.397	0.432	-0.034***	0.193	0.162	0.239	-0.076**
Yields								
Maize yield (Kg/Ha)	1,693	1,734	1,648	87				
Value of output/Ha					29,303	25,373	34,563	-9,190
Plot Manager Characteristics								
Female †	0.261	0.270	0.251	0.019	0.153	0.171	0.125	0.046***
Age (Years)	43.147	44.017	42.235	1.782***	46.817	47.508	45.759	1.749**
Education (Years)	5.028	4.961	5.098	-0.137	1.874	1.965	1.735	0.230
Household Characteristics								
Household size	4.934	4.843	5.031	-0.188***	6.476	6.490	6.455	0.035
# of HH members - [0,5]	0.981	0.964	0.999	-0.035	0.916	0.929	0.897	0.032
# of HH members - [6,14]	1.396	1.369	1.424	-0.055	1.932	1.938	1.923	0.015
# of female HH members - [15,39]	0.901	0.868	0.936	-0.068***	1.111	1.092	1.142	-0.050
# of male HH members - [15,39]	0.837	0.797	0.879	-0.081***	1.212	1.207	1.219	-0.011
# of female HH members - [40,59]	0.270	0.269	0.270	-0.001	0.386	0.388	0.384	0.004
# of male HH members - [40,59]	0.269	0.264	0.275	-0.011	0.391	0.365	0.431	-0.066**
# of HH members – 60 & above	0.280	0.311	0.247	0.063***	0.527	0.571	0.458	0.113***
Household consumption expenditures per capita	50,431	48,099	52,876	-4,777***	5,723	5,787	5,625	161
Number of plots in the holding	2.374	2.337	2.414	-0.077**	15.857	16.160	15.392	0.768
Plot Characteristics								
Owned by household †	0.904	0.927	0.879	0.048***	0.866	0.902	0.811	0.091***
Use of hired labor †	0.223	0.188	0.260	-0.071***	0.057	0.045	0.075	-0.030**
Use of organic fertilizer †	0.116	0.123	0.108	0.016**	0.183	0.246	0.088	0.157***
Use of inorganic fertilizer †	0.618	0.630	0.606	0.024**	0.404	0.419	0.383	0.036
Irrigated †	0.005	0.003	0.007	-0.004**	0.016	0.016	0.015	0.001
Soil quality good †	0.467	0.449	0.487	-0.038***	0.327	0.338	0.309	0.030
Soil quality poor †	0.113	0.112	0.114	-0.002	0.173	0.162	0.190	-0.028

Note: † denotes a dummy variable. *** p<0.01, ** p<0.05, * p<0.1. Sample of plots within a 1 kilometer radius is the comparison group for the tests of mean differences.

Table 5: OLS Imputation Model Results for Malawi - Dependent Variable: GPS-Based Plot Area (HA)

	Full Sample	Plots [d <= 1km]	Plots [d <= 500m]
Plot Area			
Farmer-reported plot area (Ha)	0.583***	0.583***	0.613***
	(0.006)	(0.007)	(0.008)
Plot Manager Characteristics			
Female †	-0.060***	-0.069***	-0.069***
	(0.013)	(0.015)	(0.016)
Age (Years)	0.003***	0.003***	0.002***
	(0.001)	(0.001)	(0.001)
Education (Years)	-0.003**	-0.004**	-0.005**
	(0.001)	(0.002)	(0.002)
Plot manager is respondent †	0.032***	0.042***	0.039***
	(0.011)	(0.013)	(0.014)
Has a chronic disease †	-0.039**	-0.043**	-0.054**
	(0.017)	(0.020)	(0.022)
Religion: Christian †	0.054**	0.062**	0.037
	(0.025)	(0.029)	(0.032)
Religion: Muslim †	-0.018	-0.005	-0.042
	(0.030)	(0.036)	(0.039)
Religion: Traditional †	0.012	0.045	-0.014
	(0.047)	(0.057)	(0.070)
Plot Characteristics			
Soil quality good †	-0.025	-0.021	-0.015
	(0.016)	(0.019)	(0.021)
Use of organic fertilizer †	0.036**	0.048***	0.033*
	(0.016)	(0.018)	(0.020)
Use of inorganic fertilizer †	0.086***	0.085***	0.064***
	(0.011)	(0.012)	(0.013)
Use of hired labor †	0.113***	0.104***	0.109***
	(0.013)	(0.015)	(0.017)
Irrigated †	-0.142**	-0.088	-0.035
	(0.072)	(0.088)	(0.111)
Household Characteristics			
# of HH members - [0,5]	0.006	0.005	0.008
	(0.006)	(0.007)	(0.007)
# of HH members - [6,14]	0.022***	0.022***	0.022***
	(0.004)	(0.005)	(0.005)
# of female HH members - [15,39]	0.011	0.001	0.001
	(0.008)	(0.009)	(0.010)
# of female HH members - [40,59]	0.057***	0.055***	0.068***
	(0.013)	(0.015)	(0.017)
# of male HH members - [15,39]	0.023***	0.022***	0.016*
	(0.006)	(0.008)	(0.008)
# of male HH members - [40,59]	0.039***	0.040**	0.044**
	(0.014)	(0.016)	(0.017)
# of HH members – 60 & above	0.030**	0.025	0.041**
	(0.015)	(0.017)	(0.018)

Table 5 (Cont'd)

	Full Sample	Plots [d <= 1km]	Plots [d <= 500m]
Household Characteristics (Cont'd)			
Wealth index	0.010***	0.008**	0.006
	(0.003)	(0.003)	(0.004)
Agriculture implement index	0.022***	0.028***	0.030***
	(0.004)	(0.005)	(0.006)
Number of plots in the holding	-0.053***	-0.052***	-0.050***
	(0.005)	(0.006)	(0.006)
Access to non-farm labor income †	-0.054***	-0.052***	-0.065***
	(0.010)	(0.012)	(0.013)
Access to non-Farm non-labor income †	-0.018*	-0.024**	-0.018
	(0.010)	(0.012)	(0.013)
Observations	18,256	13,161	9,798
Adjusted R2	0.430	0.425	0.466

Note: † denotes a dummy variable. *** p<0.01, ** p<0.05, * p<0.1. Constant, district fixed effects (30 in total) included but not reported.

Table 6: OLS Imputation Model Results for Ethiopia - Dependent Variable: GPS-Based Plot Area (HA)

	Full Sample	Plots [d <= 1km]	Plots [d <= 500m]
Plot Area			
Farmer-reported plot area (Ha)	0.871***	0.827***	0.865***
	(0.004)	(0.004)	(0.004)
Unit reported: Square Meters †	0.278***	0.270***	0.380***
	(0.021)	(0.023)	(0.027)
Unit reported: Timad†	0.241***	0.253***	0.371***
	(0.018)	(0.021)	(0.025)
Unit reported: Boy †	0.160***	0.167***	0.285***
	(0.020)	(0.022)	(0.026)
Unit reported: Senga †	0.301***	0.247***	0.336***
	(0.024)	(0.027)	(0.031)
Unit reported: Kert †	0.188***	0.200***	0.295***
	(0.028)	(0.030)	(0.034)
Plot Manager Characteristics			
Female †	-0.011*	-0.007	-0.008
	(0.007)	(0.007)	(0.007)
Age (Years)	-0.000	-0.000	0.000
	(0.000)	(0.000)	(0.000)
Education (Years)	-0.002**	-0.002***	-0.002*
	(0.001)	(0.001)	(0.001)
Religion: Orthodox †	0.023	0.037	0.031
	(0.024)	(0.025)	(0.028)
Religion: Protestant †	0.030	0.046*	0.043
	(0.025)	(0.025)	(0.029)
Religion: Muslim †	0.022	0.024	0.019
	(0.024)	(0.025)	(0.028)
Religion: Traditional †	0.028	0.045	0.039
	(0.037)	(0.039)	(0.042)
	0.030	0.046*	0.043
Plot Characteristics			
Cultivated †	0.037***	0.030***	0.028***
	(0.008)	(0.008)	(0.009)
Pasture †	0.078***	0.073***	0.066***
	(0.011)	(0.010)	(0.011)
Fallowed †	0.047***	0.047***	0.040***
	(0.013)	(0.013)	(0.015)
Soil quality good †	-0.013*	-0.007	-0.012
	(0.007)	(0.007)	(0.008)
Use of organic fertilizer †	-0.040***	-0.044***	-0.035***
	(0.008)	(0.008)	(0.009)
Use of hired labor †	0.070***	0.063***	0.051***
	(0.010)	(0.010)	(0.012)
Irrigated †	-0.022	-0.014	-0.010
	(0.015)	(0.015)	(0.016)

Table 6 (Cont'd)

	Full Sample	Plots [d <= 1km]	Plots [d <= 500m]
Household Characteristics			
# of HH members - [0,5]	0.003	0.003	0.003
	(0.003)	(0.003)	(0.003)
# of HH members - [6,14]	0.001	-0.000	-0.002
	(0.002)	(0.002)	(0.002)
# of female HH members - [15,39]	0.007**	0.005	0.004
	(0.003)	(0.003)	(0.003)
# of female HH members - [40,59]	0.012**	0.009*	0.009
	(0.005)	(0.005)	(0.006)
# of male HH members - [15,39]	0.010***	0.006**	0.006**
	(0.003)	(0.003)	(0.003)
# of male HH members - [40,59]	0.012**	0.015***	0.011*
	(0.005)	(0.005)	(0.006)
# of HH members – 60 & above	0.007**	0.006*	0.002
	(0.003)	(0.003)	(0.003)
Household consumption expenditure per capita	0.001*	0.001***	0.001***
	(0.000)	(0.000)	(0.000)
Number of plots in the holding	-0.001*	-0.001	-0.001
	(0.000)	(0.000)	(0.000)
Observations	19,966	15352	12282
Adjusted R2	0.789	0.768	0.805

Note: † denotes a dummy variable. *** p<0.01, ** p<0.05, * p<0.1. Constant, woreda fixed effects (228 in total) included but not reported.

Table 7: Tolerable Rates of Missingness in GPS-Based Plot Areas Above a Given Distance Threshold for Plot Area & Plot-Level Yield Analysis

		Plot Area	Yield
		Tolerable rate (%)	*Tolerable rate (%)*
Malawi	**1.0 km**	93 (26)	82 (23)
	500 m	52 (24)	45 (21)
Ethiopia	**1.0 km**	73 (18)	56 (13)
	500 m	48 (20)	36 (15)

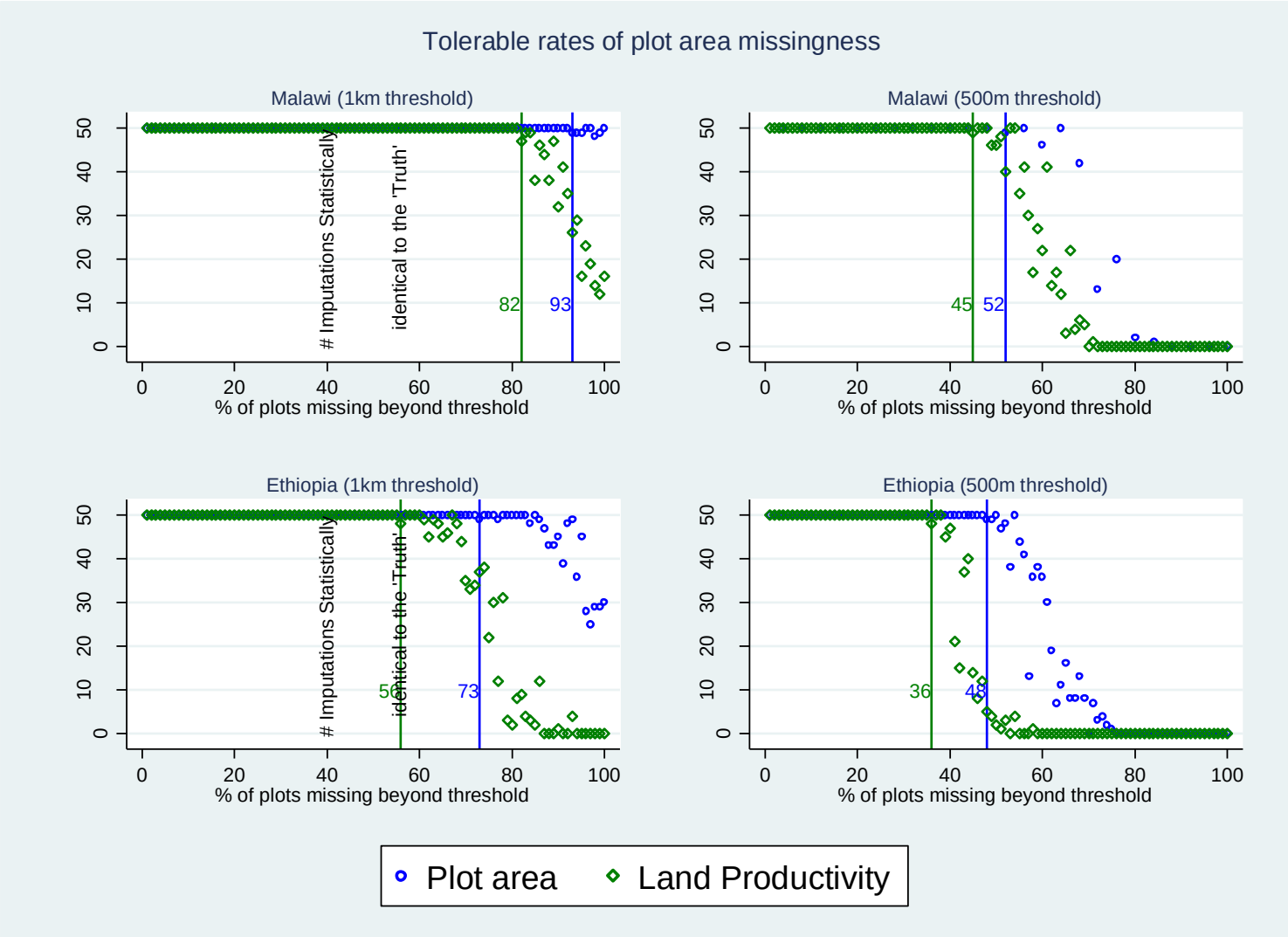
Note: The overall rates of missingness implied by the tolerable rates of missingness above a given distance threshold are noted in the parentheses.

Table 8: Country-Specific Multiply Imputed Overall Mean versus True Mean Comparisons Following Multiple Imputation At Identified Tolerable Rates of Missingness above the Distance Thresholds as Specified in Table 7

Country	Distance Threshold	Tolerable Rate of Missingness Above Distance Threshold	Variable	MI Mean	True Mean	Difference	Difference % of True Mean	RMSE	RMSE % of True Mean
Ethiopia	1 Kilometer	73	Plot Area	0.206	0.209	-0.003	-1.4%	0.003	1.4%
		56	Yield	41,141	29,303	11,839	40.4%	11,839	40.4%
	500 Meters	48	Plot Area	0.207	0.209	-0.002	-1.1%	0.002	1.1%
		36	Yield	39,628	29,303	10,325	35.2%	10,325	35.2%
Malawi	1 Kilometer	93	Plot Area	0.390	0.394	-0.004	-0.9%	0.004	0.9%
		82	Yield	1,821	1,693	128	7.5%	128	7.5%
	500 Meters	52	Plot Area	0.391	0.394	-0.003	-0.8%	0.003	0.8%
		45	Yield	1,794	1,693	101	5.9%	101	5.9%

Note: RMSE stands for Root Mean Squared Error. Plot area is in hectares. Yield is maize production in kilograms per hectare in Malawi and value of output per hectare in Ethiopia.

Figure 1: Tolerable Rates of Missingness in GPS-Based Plot Areas Above a Given Distance Threshold for Plot Area & Plot-Level Yield Analysis



6 References

Abayomi, K., Gelman, A. & Levy, M., 2008. Diagnostics for multivariate imputations. *Journal of the Royal Statistical Society: Series C (Applied Statistics),* 57(3), pp. 273-291.

Carletto, C., Gourlay, S., Murray, S. & Zezza, A., 2016. *Cheaper, faster, and more than good enough. Is GPS the new gold standard in land area measurement?,* s.l.: World Bank Policy Research Working Paper No. 7759.

Carletto, C., Gourlay, S. & Winters, P., 2015. From guesstimates to GPStimates: land area measurement and implications for agricultural analysis. *Journal of African Economies,* 24(5), pp. 593-628.

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

Dorward, A. & Chirwa, E., 2010. *A review of methods for estimating yield and production impacts,* s.l.: Centre for Development, Environment and Policy, SOAS, University of London, and Wadonda Consult 2010.

Inter-Agency and Expert Group on Sustainable Development Goal Indicators, 2016. [Online] Available at: http://unstats.un.org/unsd/statcom/47th-session/documents/2016-2-SDGs-Rev1-E.pdf [Accessed 24 6 2016].

Keita, N. & Carfagna, E., 2009. *Use of modern geo-positioning devices in agricultural censuses and surveys: Use of GPS for crop area measurement.* Durban, s.n.

Keita, N., Carfagna, E. & Mu'Ammar, G., 2010. *Issues and guidelines for the emerging use of GPS and PDAs in agricultural statistics in developing countries.* Kampala, Uganda, s.n.

Kilic, T., Palacios-López, A. & Goldstein, M., 2015. Caught in a Productivity Trap: A Distributional Perspective on Gender Differences in Malawian Agriculture. *World Development,* Volume 70, pp. 416-463.

Kilic, T., Zezza, A., Carletto, C. & Savastano, S., forthcoming. Missing(ness) in action : selectivity bias in GPS-based land area measurements. *World Development.*

Marchenko, Y. V. & Eddings, W., 2011. *A note on how to perform multiple-imputation diagnostics in Stata,* College Station, TX: StataCorp.

Rubin, D. B., 1987. *Multiple Imputation for Nonresponse in Surveys.* Hoboken, NJ: Jon Wiley & Sons.

Rubin, D. B., 1996. Multiple Imputation After 18+ Years. *Journal of the American Statistical Association,* 91(434), pp. 473-489.

van Buuren, S., Boshuizen, H. C. & Knook, D. L., 1999. Multiple imputation of missing blood pressure covariates in survival analysis. *Statistics in Medicine,* 18(6), pp. 681-694.