Holden, Stein Terje; Makate, Clifton; Tione, Sarah

Working Paper

Missing parcels and farm size measurement error: Do nationally representative surveys provide reliable estimates?

Centre for Land Tenure Studies Working Paper, No. 04/25

Provided in Cooperation with:
Centre for Land Tenure Studies (CLTS), Norwegian University of Life Sciences (NMBU)

Norwegian University of Life Sciences (NMBU)

Missing Parcels and Farm Size Measurement Error: Do Nationally Representative Surveys Provide Reliable Estimates?

Stein T. Holden, Clifton Makate and Sarah Tione

Norwegian University of Life Sciences
Centre for Land Tenure Studies

Centre for Land Tenure Studies Working Paper 04/25

ISBN: 978-82-7490-334-0

Missing Parcels and Farm Size Measurement Error: -Do Nationally Representative Surveys Provide Reliable Estimates?

By

Stein T. Holden[1*], Clifton Makate[1,2], and Sarah Tione[3]

[1]School of Economics and Business, Norwegian University of Life Sciences (NMBU), Box 5003, 1432 Ås, Norway. [*]Corresponding author Email: stein.holden@nmbu.no

[2]Norwegian Geotechnical Institute (NGI), P.O. Box. 3930 Ullevål Stadion, N-0806 Oslo, Norway.

[3]Department of Agricultural and Applied Economics, Lilongwe University of Agriculture and Natural Resources, Bunda Campus, Lilongwe, Malawi.

Abstract

We assess the reliability of measured farm sizes (ownership holdings) in the Living Standard Measurement Study – Integrated Surveys on Agriculture (LSMS-ISA) in Ethiopia and Malawi based on three survey rounds (2012, 2014, 2016) in Ethiopia and four rounds (2010, 2013, 2016, 2019) in Malawi. Using the balanced panel of households that participated in all the rounds, we utilized the within-household variation in reported and measured ownership holdings that were mostly measured with GPSs and/or rope and compass. While this gives reliable measures of reported holdings, we detect substantial under-reporting of parcels over time within households that largely have been overlooked in previous studies. The problem causes an unrecognized bias in agricultural statistics. We find that the estimated farm sizes within survey rounds are substantially downward biased due to systematic and stochastic under-reporting of parcels. Such biases are substantial in the data from both countries, in all survey rounds, and in all regions of each country. We estimate models with alternative estimators for the ownership holding share of maximum within-household holding to examine factors associated with variation in reported farm sizes. Based on the analyses, we propose that the maximum within-household reported farm sizes over several survey rounds provide a more reliable proxy for the real farm size, as these maximum sizes are less likely to be biased due to parcel attrition. The ignorance of *this non-classical measurement error is associated with a downward bias of 12-41% in average and median farm sizes and an upward bias in the Gini coefficients for farm size distributions. We propose ideas for follow-up research and improvements in collecting these data types and draw relevant policy implications.*

Key words: Farm size measurement, plot attrition, measurement error, LSMS-ISA, Ethiopia, Malawi.

JEL codes: C81, C83, Q12, Q15.

1. Introduction

Good agricultural statistics are essential for planning and dealing with many global challenges associated with climate change and global, national, and local food security (Carletto 2021). Carletto et al. (2021) argue for the importance of renewed attention to data quality issues for advancing the research frontier in agricultural economics and designing better agricultural policy. Developing countries that rely on agriculture as a primary source of livelihood for a large share of their population are among the most vulnerable to climate change (Lowder et al. 2016). Recent conflicts have further contributed to instability in global prices for food and energy and have enhanced global food insecurity. Rural transformation, rural-urban, and international migration are putting more pressure on areas on the receiving end. While economic development creates new opportunities in rural transformation processes, climate shocks and social unrest are among the push factors associated with more desperate migration. Agricultural development and intensification are essential to reduce the extent of desperate migration and enhance food security. Good agricultural policies are crucial, and good agricultural statistics are relied on to tackle these challenges and promote sustainable agricultural intensification (World Bank 2021).

The 2008 World Development Report on Agriculture for Development became a vital driver in generating better agricultural statistics as a basis for a new push for agricultural development. One outcome was nationally representative household farm surveys such as the Living Standard Measurement Study - Integrated Surveys on Agriculture (LSMS-ISA). They provided essential data for analyzing important policy issues in developing countries. Modern technologies such as handheld GPSs and preprogrammed tablets linked to cloud servers have reduced costs and improved the quality of such survey data. Accompanied by improved methods for area measurement, the role of potential measurement error and its implications for various types of estimation purposes and data reliability has become a new area of research (Carletto et al. 2013; 2017; Abay et al. 2019; 2021; 2023a: 2023b; Burke et al. 2019; Gourlay et al. 2019; Kilic et al. 2017a; b; Wossen et al. 2022). Handheld GPSs provide much more reliable estimates of farm parcel sizes than farmers' estimates of parcel and farm sizes, which were often used in the past (Carletto et al. 2013; 2017). Self-reported area data include systematic biases that depend on the parcel size, rounding errors, and influences from local measurement units (Abay et al. 2019; Carletto et al. 2017; Holden and Fisher 2013). Such systematic errors could affect yield estimates and explain the frequently found phenomenon of an inverse relationship between parcel yield and parcel size, as the parcel size is used to construct the yield variable. Such measurement errors could also systematically affect measures aggregated to the farm level, where a farm may contain a varying number of parcels (Holden and Fisher 2013). Kilic et al. (2017a) used multiple imputations (MI) methods to predict more accurate area measures of unmeasured parcels. In a follow-up study, Kilic et al. (2017b) did a more comprehensive test of the MI approach with the 2013/2014 Ethiopia Socioeconomic Survey Wave II data from Ethiopia and the Integrated Household Survey 2010/2011 (IHS3) for Malawi where they had more complete parcel-level data measured with GPS and farmers' own estimated areas[i]. They found that the MI approach, to a large extent, could correct biases associated with incomplete coverage with GPS-measurement of parcels when farmers' estimated parcel sizes were available.

It is not possible to obtain the true measure of farm size. Areas measured with handheld GPS are also measured with error; the relative error size is inversely related to parcel size. However, unlike self-reported area sizes, GPS-based average parcel size estimates are not found to be biased even for tiny parcels (Carletto et al. 2017). Average farm sizes aggregated from several parcels estimated by GPS are, therefore, also not likely to have a systematic bias given that all the parcels on the farm were measured with GPS before aggregation (assuming the error is uncorrelated across parcels).

Our study focuses on a different missingness problem that we elaborate on below, which is revealed only when multiple survey rounds for the same households (balanced panel) are combined. We investigate how unreported or missing parcels can influence farm size measurement and its reliability in the LSMS-ISA data from two countries, Ethiopia and Malawi. We aim to assess the farm sizes and potential measurement errors in these data over time, where GPSs are the primary device in measuring farm parcels. We ask the question: Can these extensive surveys provide reliable estimates of farm size changes over time? If yes, the data may also be used to provide reasonably reliable estimates of farm size distributions such as Gini-coefficients and cumulative farm size graphs and how these measures change over time within the smallholder sector in these countries[ii]. To assess the reliability of such estimates, we utilize only farm households that are repeatedly surveyed in each country for three panel rounds in Ethiopia and four panel rounds in Malawi. We propose that there is a high probability of under-reporting of land parcels due to the drudgery of reporting data from and measuring parcels. In this study, we assess the extent of such potential parcel attrition and separate it from real within-household farm size change over time that can occur due to inheritance and bequeath, land purchases and sales, administrative redistributions, and land grabs.

We use censored Tobit models to estimate the reported farm sizes as shares of maximum within-household farm sizes across survey rounds. We also test alternative estimators. These models provide insights about possible real farm size changes but, more importantly, strong indications of widespread stochastic under-reporting of plots. To our knowledge, this is the first study to provide such comprehensive evidence. This is the main contribution of our study. We find that these biases in estimated farm sizes due to the wicked plot attrition problem cannot easily be overcome with econometric estimators that attempt to control for real farm size changes and parcel or plot attrition with plot count indicators, although the models help to scrutinize the evidence.

We conclude that the maximum reported within-household farm size over repeated survey rounds represents the most reliable measure of household farm size and is the least likely to suffer from downward bias due to plot attrition. We compare the farm size distributions based on these reported maximum within-household farm sizes with the reported farm sizes in each survey round in Ethiopia and Malawi. These comparisons demonstrate substantial downward biases in the range of 12-29% for mean farm sizes and in the range of 30-39% for median farm sizes in Ethiopia, and in the range of 30-39% in mean and of 30-41% in median farm sizes in Malawi.

Our study reveals a type of measurement error that largely has gone under the radar and has received too little focus until now. Any studies utilizing the same data that have attempted to study land productivity and associated it with farm size based on these data should be revisited with these new insights in mind. It is highly likely that this parcel or plot attrition is also associated with under-reporting of parcel or sub-parcel (plot) output and possibly input use if such reporting is done at the plot level and not at the parcel or household level.

In part 2, we outline a theoretical framework for the study, followed by a description of the data management strategy in part 3. In part 4, we present the main findings for Ethiopia and Malawi. In part 5, we discuss the results before we conclude.

2. Theoretical framework: Explaining observed farm size variation due to real changes and measurement error

2.1. Factors that explain real farm size variation

The main reasons that can explain the within-household changes in farm sizes over time are the following;

a) *Inheritance and bequeath of land within families.* Young household heads are more likely to inherit land, and old household heads with adult children are likelier to bequeath their land to the next generation. Changes in heads of households may also be associated with such changes in ownership holding size, e.g., related to divorce or marriage and takeover of farms.
b) *Purchases or sales of land.* In countries with active land sales markets, farms may change owners, but there could also be changes in farm sizes associated with sales or purchases of parcels of land. Such markets tend to be thin in developing countries and are not likely to influence farm size changes for a large share of a random sample of household farms.

c) *Administrative expropriations and land redistributions.* Depending on national land policies, such administrative redistributions are more common in some countries than others. Such redistributions may also be more common during rapid urban expansion and transformation in peri-urban areas.

d) *Private land takings and losses.* The extent to which such processes are common depends on land abundance, national policies, and enforcement capacity/tenure security. Such events as sudden shocks may affect rural households.

2.2. Theories to explain errors in farm size measurement

We need theories to explain potential non-classical measurement errors that can lead to systematic biases in reported and estimated farm sizes. These theories should help explain the under-reporting of ownership holdings and possible mistakes in reported holdings. Recent literature distinguished measurement errors due to misreporting and misperceptions (Abay et al. 2021; 2023a: 2023b: Wossen et al. 2022). Our basic assumption is that the introduction of GPS or other high-quality measurement of farm parcels/plots eliminates most errors associated with misperceptions that can lead to errors in measured plot, parcel, and farm sizes. However, this important quality improvement does not prevent errors due to misreporting of plots/parcels.

Conditional on finding such an unexplained gap that the standard theories above cannot explain, we suggest a set of propositions based on theories in new institutional economics, such as imperfect information and transaction cost theories. Information asymmetries and the high costs of obtaining information may contribute to explaining that a substantial non-classical measurement error due to misreporting exists in household ownership holdings. These propositions are as follows:

Prop.1: *Farmers have incentives to hide some of their parcels to reduce the burden of answering all questions in the survey*.

Prop.2. *Enumerators also have incentives to reduce the number of parcels recorded for each household to reduce their work burden*. Prop. 1 and 2 may also imply that farmers and enumerators collude to reduce their joint burden associated with the data collection. The extent of enumerators' supervision, transparency, and motivation may vary over survey rounds and possibly across data collection teams, which may cause spatial and intertemporal variations in data quality.

Prop.3. *Surveys tend to focus only on the main (large)nearby parcels of a farm and leave out small parcels of less significance and parcels that are located far away*. Survey budgets and standards may be set that cause less than complete parcel measurement.

Prop.4. *Rented-out parcels are more likely to be left out from the survey as such parcels are not managed by the household included in the survey*. The owner may be unable to provide much production data from rented-out parcels.

Prop.5. *Improvements in the data collection technologies and methods have reduced information asymmetries and transaction costs over time.* The new CAPI tools have also made it easier to monitor enumerators, and the information and transaction costs associated with data collection have been substantially reduced. These technological improvements may imply that parcel attrition has been reduced over time. Over time, such a reduction in parcel attrition may have led to an artificial increase in farm sizes in balanced household farm panels.

We are unable to test each of these theoretical propositions explicitly. In this study, we only aim to assess the size of the measurement error problem by estimating farm size variation and controlling for real farm size changes to the extent that suitable control variables exist in the publicly available data. A challenge we face is that we do not know the true farm size for each household in the LSMS-ISA data. It is, therefore, difficult to identify a proper benchmark to get a reliable estimate of the measurement error. Consequently, we attempt a second-best approach to assessing the significance of such errors. We use the largest within-household ownership holding size identified over the three or four survey rounds as the benchmark. We do this as we believe that the main problem in the data is missing (omitted) parcels/plots, and the largest within-household estimate of the farm size (ownership holding) is, therefore, least likely to suffer from this omitted parcel problem. This theoretical framework is the basis for our data management and estimation strategy, which we outline in the next section.

3. Data management and estimation strategy

We create the three (Ethiopia) and four (Malawi) rounds of balanced household (households for which there are data for all these rounds) LSMS data with all land size relevant variables such as the size of all parcels (GPS-measured), number of parcels, number of owned parcels, number of operated parcels, number of rented out parcels, number of rented in parcels, number of borrowed in/out parcels, inherited and bequeathed parcels between survey rounds, expropriated parcels, and parcels received through redistribution.

One challenge is to match parcels from survey round to survey round within households. Parcels should be stable over time and easier to match than plots, which could be sub-units of parcels that may change with crop planting patterns. We will return to this issue later.

Between survey rounds, we make the following within-household identity for ownership holding based on GPS-measured parcels (as far as possible) based on data from survey rounds *t*1 and *t*2:

$$(1)\quad A_{t2}^{o} = A_{t1}^{o} + A_{t1-t2}^{p} + A_{t1-t2}^{i} - A_{t1-t2}^{s} - A_{t1-t2}^{b} - A_{t1-t2}^{e} + \Delta A_{t1-t2}^{M}$$

where superscripts *o* represents ownership holding measured at times *t*1 and *t*2, *p* represents purchased holding, *i* inherited holding, *s* sold holding, *b* bequeathed holding, and *e* expropriated holding, where these changes had happened between period *t*1 and *t*2 when ownership holdings were measured. The identified discrepancy ΔA_{t1-t2}^{M} represents the unexplained changes due to measurement error/data gap. A similar identity can be set up for *t*2 versus *t*3 to identify a similar household-level discrepancy in farm size determination.

It is also possible that the operational holding (area being farmed by the household in a specific year) deviates from the ownership holding because the household rents in or rents out land and because all the owned land may not be farmed but is left fallow, and some land may be lent out or lent in.

Unfortunately, the LSMS-ISA survey data do not provide complete information regarding the components in the identity in equation (1). In the survey, there is a question about the origin of each parcel of land but not when it has been received such that it is possible to verify with certainty whether it has changed since the previous survey rounds, which typically took place two to three years earlier (in rounds two, three and four of the three- and four-rounds balanced household panels).

While GPS positions exist for the parcels, these are not publicly available because of the need to protect the anonymity of the households. We recommend that those with access to these GPS records do a parcel-level matching of the data for the balanced panel to verify parcel attrition over time more accurately based on exact location records. However, such parcel matching is beyond what we can do based on the publicly available data. We hope our study draws more attention to the importance of such across-round data verification to reduce parcel attrition and improve data quality.

Based on the data, we have attempted to approximate the ownership holding of households by survey round. We have subtracted rented and borrowed land from the declared parcels to approximate the ownership holding sizes based on the measured areas. We acknowledge that this may represent an underestimation of ownership holding as it is possible that some owned parcels have not been reported, e.g., because they are rented out or, for other reasons, have not been declared. Such attrition is more likely to be detected with repeated survey rounds of the same households. We have constructed two new within-household variables for ownership holdings to assess the extent of such possible attrition. The first is the time-invariant maximum within-household ownership holding. The second is the time-variant ownership holding as a share of the maximum within-household ownership holding in each survey round.

Our study is exploratory in the sense that we want to measure the relative size and variation in this measurement gap. The reference point is the maximum within-household measured ownership holding observed over the three or four rounds.

We then measure each household’s ownership holding in each survey round as a share of its’ maximum measured ownership holdings over the three (Ethiopia) and four (Malawi) survey rounds. We then explore the variation in this ownership share of the maximum size across households and survey rounds. We estimate how much of the share of the maximum holding size is influenced by the real changes in farm sizes by including control variables associated with such real changes in the form of inheritances, bequeaths, sales and purchases, and administrative redistributions and land takings. We attribute the residual deviation from maximum farm size to the imperfection information theories that explain parcel attrition. We use several available control variables for this. First, we include variables associated with land being rented out as such land is more likely to have been unreported. Second, we use parcel counts in each survey round and within-household deviation in parcel counts over survey rounds. As the division into sub-parcels[iii] may change from survey round to survey round and depend on the cropping pattern, using such sub-parcel counts is not a waterproof measure of parcel attrition. Still, it can nevertheless be a good indicator. We expect that a higher parcel count, on average, is associated with a larger and more complete measure of the farm size.

As proposed by our theoretical framework, we take the maximum ownership holding over time as the reference point, as this area is the least likely to suffer from attrition.

We tailor the approach to the specific contextual and policy situations and survey instruments used in the two countries.

Hypothesis: *Parcel (plot) attrition varies stochastically across survey rounds and causes substantial within-household measurement error and downward bias in measured within-round average farm sizes.* We construct the dependent variable as a share of the within-household maximum farm size across survey rounds as the benchmark.

We need complementary strategies to investigate the reasons for the variation in the size of within-household farms over time. In addition to the real area changes that we introduce controls for, we add the following variables as tests and controls for within-household stochastic attrition:

a. Total number of (sub-)parcels reported in the survey round
b. Deviation in the maximum number of (sub-)parcels reported in the survey round compared to the round with the highest number of (sub-)parcels reported.
c. Number of unmeasured (sub-)parcels in the survey round.

Given that this attrition is stochastic, we assume that a larger parcel count positively correlates with the measured farm size and the ownership share of maximum holding in a given survey round. Furthermore, we think a larger deviation from the maximum within-household number of (sub-)parcels is associated with a smaller farm size estimate relative to the maximum within-household farm size (lower ownership share).

The larger the number of unmeasured (sub-)parcels, the smaller the measured farm size is assumed to be. The number of unmeasured but reported parcels was low in Ethiopia and Malawi (1.3-2.1% of the parcels) balanced panel data. We control for it in our estimation of relative farm sizes in the form of ownership holding shares of max within-household ownership holdings.

We reduce possible outlier errors in the estimated data by winsorizing the measured farm sizes at a 1% level at each distribution end. We also assessed the effect of varying degrees of winsorizing the data. We found that doing this at a 1% level was sufficient to remove random noise that could affect the tails in the distribution and make maximum and minimum area measurements unreliable.

A small share of the households dropped out after the initial survey round. To correct for possible attrition bias due to the drop-out of households, we ran models for the data from the first survey round with the attrition dummy as a dependent variable and household/farm characteristics as explanatory variables. We constructed attrition weights based on the ratio between the predicted attrition rates with all variables included and a model where we only included the insignificant variables. We use these inverse probability weights in weighted regressions to test for and correct this type of possible attrition bias in the data.

We estimate the ownership shares using censored Tobit models censored from above at one for each survey round and jointly as a panel for all rounds within each country. These models allow us to assess and separate the relative farm size changes associated with the variables that control for real changes in ownership holding shares over time from changes related to incomplete reporting of areas.

Censored Tobit models are sensitive to non-normality and heteroskedasticity. As a robustness check, we therefore also estimated alternative models in the form of fractional probit models, symmetrically censored least squares estimator (SCLS), and panel stochastic frontier models for the ownership holding shares of max holding share as these are all in the zero-one range. We compare the parameters across models and also the cumulative predicted outcome distributions.

Furthermore, we compare actual and predicted ownership share distributions across alternative models. We compute the farm size distributions in each survey round against the across-year maximum household farm size distribution as indicators of the potential bias in ownership holding sizes based on data from each round to assess the extent of bias in the farm size in each survey round caused by within-household stochastic parcel/plot attrition bias. We also consider the spatial and inter-temporal variation in ownership holding shares to evaluate whether that can provide insights about variation in the survey quality in reducing the extent of this type of measurement error. Finally, we also generate Gini-coefficients for the measured ownership holdings and the maximum holdings to assess whether parcel attrition is associated with bias in estimates of inequality in land holding distributions.

Parcel-level GPS coordinates are not publicly available, which makes it challenging to generate balanced household-parcel-level panel data to scrutinize the (sub-)parcel attrition in more detail. The enumerator instructions specified that enumerators should use parcel IDs from the previous round to facilitate such matching of parcels over time. However, careful data inspection revealed that this had not been rigorously implemented. We demonstrate this by comparing selected households at p10, p25, and p50 of the ownership share distributions in each survey round in terms of their parcel IDs across years, see Appendix 4. We also created a variable in the form of the average ownership holding share across survey rounds and did the same exercise for households at p10, p25, p50, p75, and p90 in each country. We expect to find a better match of parcel sizes by parcel IDs across years for the p75 and p90 households. We present the detailed parcel distributions by year and measured parcel sizes in Appendix 4.

4. Results

Below, we present the detailed data analysis for each country by first looking at some basic descriptive results, then by running censored Tobit models for ownership holding shares by survey round and as panel models to investigate alternative explanations for the within-household farm size variation across survey rounds, and to assess the severity of stochastic plot attrition, its implications for estimated farm sizes, and cumulative farm size distributions. We demonstrate that stochastic plot attrition results in large non-classical measurement errors in farm sizes that are severely downward biased in each survey round.

4.1. Ethiopia

We have used the 2012, 2014, and 2016 survey rounds for Ethiopia. Table 1 presents the balanced household sample by region in Ethiopia.

The size of the deviation from the maximum holding (measured as a share of the maximum holding) is an indicator of the extent of within-household reported change in ownership holding over time. It may be due to inheritance, bequeath, administrative allocation, purchase and sale (rare in Ethiopia), or parcel/plot attrition. Over the short period from 2012 to 2016, we expect the extent of inheritances, bequeaths, and administrative allocations to be pretty small and land sales to be non-existent since land sales are illegal in Ethiopia. A large deviation may indicate substantial attrition (parcels omitted in surveys). Based on our theoretical framework, we have included land renting for several reasons: a) households are more likely to report parcels they cultivate themselves, b) the land rental market (including sharecropping) is very active in Ethiopia, c) households in Ethiopia may not perceive sharecropping as a form of renting, d) rented (sharecropped) out plots are less likely to be reported. The literature shows that households without oxen and female-headed households are more likely to rent out their land (Holden et al. 2008; 2011).

Table 1. The balanced panel for Ethiopia, distribution of households by region and year

	Year of survey			
Region	2012	2014	2016	Total
Tigray	297	297	297	891
Afar	76	76	76	228
Amhara	572	572	572	1,716
Oromiya	563	563	563	1,689
Somalie	168	168	168	504
Benishangul Gumuz	96	96	96	288
snnp	776	776	776	2,328
Gambella	85	85	85	255
Harari	97	97	97	291
Dire Dawa	100	100	100	300
Total	2,830	2,830	2,830	8,490

Table 2 presents basic statistics for reported and measured ownership holding sizes by survey round where these estimates are un-winsorized or winsorized at the 1% level. These estimates are then compared with winsorized maximum within-household ownership holding sizes over the three survey rounds, where we alternatively have winsorized the maximum ownership holdings at 1, 2, and 5% levels. We see an astonishing gap between the estimates for each survey round and the maximum holding sizes over all three survey rounds. Winsorizing the annual data at 1% creates a downward trend over the years in mean and median ownership holding sizes, as would be expected due to population growth and bequeaths of land from parents to children. We, therefore, think the data quality may have been improved with this adjustment of outlier observations.

Table 2. Estimated ownership holdings in ha based on a three round balanced sample of household-parcel panel data from Ethiopia

	Unwinsorized Ownership holdings, ha			1% winsorized Ownership holdings, ha			Max ownership holdings 2012-2016 in ha, winsorized at:		
	2012	2014	2016	2012	2014	2016	1%	2%	5%
Mean	1.070	1.317	1.141	0.997	1.102	1.066	1.499	1.435	1.317
Median	0.645	0.731	0.695	0.645	0.731	0.695	1.050	1.050	1.050
P25	0.218	0.271	0.250	0.218	0.271	0.250	0.486	0.486	0.486
P75	1.343	1.468	1.372	1.343	1.468	1.372	1.948	1.948	1.948
P90	2.306	2.497	2.506	2.306	2.497	2.506	3.268	3.268	3.268
sdev	1.952	4.535	2.100	1.165	1.254	1.238	1.506	1.300	1.019
n	2830	2830	2830	2830	2830	2830	2830	2830	2830
Gini	0.560	0.605	0.562	0.528	0.529	0.531	0.493	0.472	0.434

To better understand the potential upper bound of the attrition frequency across the three survey rounds within households, we construct a new variable: the household and year-specific ownership holding divided by the maximum within-household ownership holding over the three survey rounds. We graph the ownership share distributions of maximum within-household ownership holding distribution for each survey round, and we know that one of the three rounds is represented with the maximum ownership holding in the three panel-years. To assess whether random measurement errors cause outliers, especially in maximum ownership holding, we compare the unwinsorized data with alternative winsorized data at 1, 2, and 5% on both sides of the distributions. We present cumulative density distributions for each survey round with the winsorized data as overlays in each panel year in Figure 1.

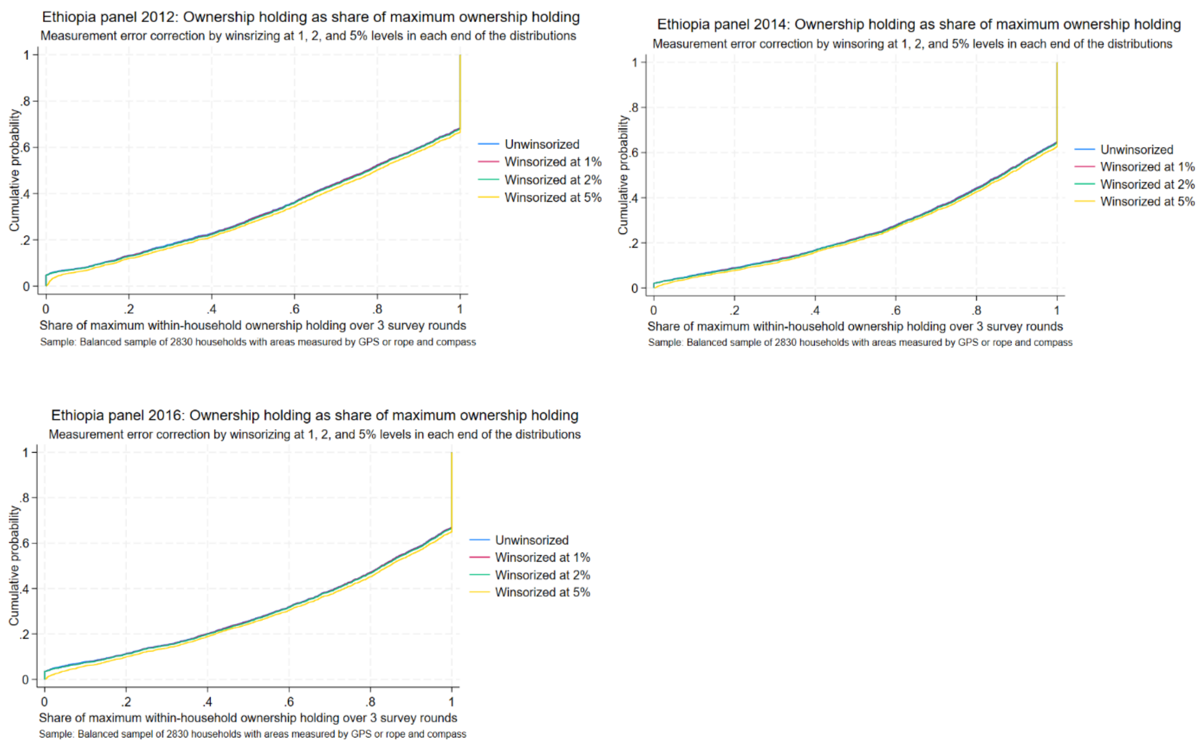

Figure 1. Within-household ownership holding shares of maximum household ownership holdings over three survey rounds in Ethiopia with alternative levels of winsorizing.

Figure 1 shows that a slightly larger share of households were at the maximum farm size in 2014 and 2016 than in 2012. This may indicate a higher level of parcel attrition in 2012, which is also consistent with the fact that the total number of reported parcels/plots was lower in 2012. Another important insight from Figure 1 is that the measurement error corrections by winsorizing data at 1, 2, and 5% levels had a minimal effect on the cumulative ownership share distributions. This minimal effect on the distributions indicates that only a tiny part of the variation in within-household ownership shares is due to random measurement errors. However, the graphs do not tell how much of the deviations from one in ownership shares are due to changes over time in inheritances, bequeaths, purchases, sales, or administrative redistributions, and within-survey round attrition of parcels. We use econometric methods to explore these deviations. While we have data on the origin of the parcels, we do not know when the parcels were received or whether parcel transfers occurred within the panel period (2012-2016). Figure 1 indicates surprisingly large changes in ownership holding sizes compared to the maximum holding size over this fairly limited period from 2012 to 2016. For example, we see in 2012 that 20% of the sample households had an ownership holding that was less than 40% of the maximum ownership holding size within the 2012-2016 period. For 2014 and 2016, about 20% of the sample had ownership holding sizes below 50% of the maximum holding size over the 2012-2016 period. Knowing that land sales are illegal in Ethiopia and that administrative redistributions have become much less common than before makes it hard to believe that inheritances and bequeaths have resulted in large farm size changes over such a limited time period.

To inspect the importance of parcel attrition, we include the number of (sub-)parcels (plots) measured in each survey round and the within-household deviation in number of parcels from the maximum number across survey rounds, see Figure 2a and 2b.

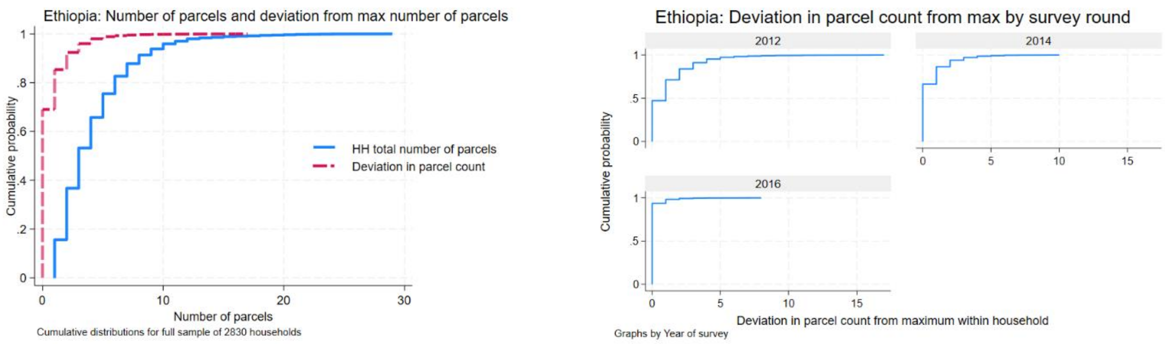

Figure 2a and Figure 2b. Ethiopia: Parcel count and deviation from max parcel count (total and by survey round) for households

Figure 2a shows that more than 30% of the households have reported a varying number of sub-parcels (plots) across survey rounds. This represents no absolute evidence of parcel attrition, as parcels may have been divided into variable plots depending on changing cropping patterns over the years. However, Figure 2b shows that the number of reported parcels was systematically lower in 2012 than in the two following rounds. This may be associated with the smaller average farm size in 2012 in the un-winsorized data in Table 2.

To test our hypothesis that missing (omitted) parcels cause under-reporting of farm sizes, we have estimated censored Tobit models separately for each survey round and jointly with a panel Tobit model. The complete model results are presented in Table 3. For robustness assessment, we estimated fractional probit, symmetrically censored least squares estimator (SCLS), and panel stochastic frontier models.

Appendix A2 presents the results. These models provide similar results as the censored Tobit models, and the choice of estimator did not give reasons to change any of our interpretations of the main results.

For the 2012 survey round, the intercept share for a male-headed household with more than two oxen, less than 31 years old, and located in the Tigray region is 0.75 of the maximum own holding size. For household heads that are above 60 years the holding size is 11.7 percentage points higher. This change may represent the effect of bequeath on ownership holding and indicate that the youngest household head group aged <31 years may later have, *ceteris paribus*, gained these percentage points in relative farm size compared to the oldest group. This change only captures a 11.7/25 share of the gap in the average ownership holding share of maximum holding. The logic behind this is that the oldest household heads were closer to their maximum holding in 2012 than the youngest household heads, who were more likely to inherit land during the 2012-2016 period. Two of the variables associated with a higher likelihood of renting out land (female-headed households, households having no oxen, or only one ox for land cultivation) had significantly smaller ownership holding shares than other households. This indicates that those who rent out land tend to under-report such parcels, leading to the type of bias we hypothesized to find. The administrative redistributions or land-taking indicators were insignificant and had high standard errors. The total plot count was highly significant (at 0.1% level) and positively correlated with the ownership holding share, indicating that higher counts are associated with less likelihood of attrition. The deviation from the maximum plot count was highly significant (at 0.1% level) and had a negative sign. We interpret this as robust evidence of plot attrition explaining low ownership shares. One less sub-parcel (plot) counted in 2012 than the maximum count is associated with a 3.4 percentage point lower ownership share measured, and about 50% of the sample may have under-reported the number of plots in 2012 (Figure 2b).

Furthermore, one unmeasured plot/parcel is significantly (at a 0.1% level) associated with a 4.0 percentage point smaller ownership share measured.

Table 3. Ethiopia censored Tobit models for ownership shares of maximum holdings split by survey round and pooled panel censored Tobit model

	Tobit12	Tobit14	Tobit16	Tobit1216
Year(s)→	2012	2014	2016	2012-2016
Female-headed hh	0.018	0.028	0.006	0.013
	(0.03)	(0.03)	(0.03)	(0.01)
Two oxen(base)	0.000	0.000	0.000	0.000
No ox	-0.071**	0.010	-0.076**	-0.044****
	(0.03)	(0.03)	(0.03)	(0.01)
One ox	-0.070**	0.062**	0.000	-0.002
	(0.03)	(0.03)	(0.04)	(0.01)
Three-plus oxen	0.052	0.033	0.151***	0.084****
	(0.05)	(0.05)	(0.05)	(0.01)
Household size	-0.027	0.007	-0.011	-0.003
	(0.02)	(0.01)	(0.01)	0.00
Tot. Labor units	0.048**	0.008	0.020	0.013**
	(0.02)	(0.02)	(0.01)	(0.01)
Age oldest child	0.003	0.000	0.000	0.000
	(0.00)	(0.00)	(0.00)	(0.00)
Base: Age hhh 20-30	0.000	0.000	0.000	0.000
Age hhh 31-40	0.062*	-0.034	0.042	0.022
	(0.04)	(0.04)	(0.05)	(0.01)
Age hhh 41-50	0.032	-0.011	0.025	0.020
	(0.04)	(0.05)	(0.06)	(0.02)
Age hhh 51-60	0.104**	-0.045	0.008	0.025
	(0.04)	(0.05)	(0.06)	(0.02)
Age hhh >60	0.117***	0.040	-0.033	0.041**
	(0.04)	(0.05)	(0.06)	(0.02)
Involuntary loss of farm	0.052	-0.039	-0.119	-0.046
	(0.11)	(0.09)	(0.11)	(0.07)
Displacement by Gov.	-0.060	-0.406**	-0.282	-0.236***
	(0.20)	(0.16)	(0.34)	(0.08)
Local unrest shock	-0.168	0.042	0.121	0.044
	(0.13)	(0.17)	(0.08)	(0.05)
Sqrt(Distance to admin center)	-0.003	-0.002	-0.006**	-0.003***
	(0.00)	(0.00)	(0.00)	(0.00)
Total plot count	0.015****	0.015****	0.015****	0.015****
	(0.00)	(0.00)	(0.00)	(0.00)
Deviation from max plot count	-0.034****	-0.040****	-0.064****	-0.036****
	(0.01)	(0.01)	(0.02)	(0.00)
Number unmeasured parcels	-0.040****	-0.110****	-0.210	-0.048****
	(0.01)	(0.02)	(0.26)	(0.01)
Oromia region (base)	0.000	0.000	0.000	0.000
Tigray	-0.034	-0.041	0.038	-0.009
	(0.05)	(0.05)	(0.06)	(0.02)
Afar	-0.180**	-0.143**	0.029	-0.094***
	(0.07)	(0.07)	(0.07)	(0.03)

Amhara	0.014	-0.012	-0.012	-0.001
	(0.04)	(0.04)	(0.04)	(0.01)
Somalie	-0.091*	0.005	0.104**	0.002
	(0.05)	(0.05)	(0.05)	(0.02)
Beninshangul Gumuz	0.004	0.021	-0.106*	-0.032
	(0.06)	(0.06)	(0.06)	(0.03)
SNNP	0.010	0.052	0.031	0.037***
	(0.03)	(0.04)	(0.04)	(0.01)
Gambella	-0.262****	0.022	-0.006	-0.075***
	(0.06)	(0.06)	(0.06)	(0.03)
Harari	0.095	-0.194**	-0.007	-0.035
	(0.08)	(0.09)	(0.11)	(0.03)
Dire Dawa	0.020	0.035	0.063	0.049*
	(0.06)	(0.06)	(0.06)	(0.03)
2012 base year				0.000
2014 panel year				-0.085****
				(0.01)
2016 panel year				-0.150****
				(0.01)
Constant	0.750****	0.685****	0.741****	0.802****
	(0.08)	(0.08)	(0.09)	(0.03)
var(e.own~1)	0.172****	0.151****	0.167****	
	(0.01)	(0.01)	(0.01)	
sigma_u				0.000
				(0.01)
sigma_e				0.409****
				(0.00)
Prob > chi2	0.0000	0.0000	0.0000	0.0000
Number of observations	2830	2830	2830	8490

Note: Models corrected for attrition bias with IPW. Robust standard errors in parentheses. Significance levels: *: 10% level, **: 5% level, ***: 1% level, ****: 0.1% level.

There were some regional differences, with ownership shares being significantly smaller in the Afar and Gambella regions compared to the Oromia region used as a base.

The key plot attrition variables are similarly highly significant, with similar coefficient sizes in the 2014, 2016, and panel models in Table 3, further supporting our primary hypothesis that plot attrition contributes to the under-reporting of farm sizes.

The pooled censored Tobit model results for all three survey rounds are also presented in Figure 3, and the last model in Table 3. The variation in the confidence intervals in Figure 3 shows the variation in the precision of the estimates of the different variables. This model allowed us to test for differences across survey rounds with panel-round fixed effects. The dummies for the 2014 and 2016 survey rounds are highly significant (at 0.1% levels) and indicate that the ownership shares were 8.5% lower in 2014 and 15% lower in 2016 than in 2012. The pooled model indicates that land renting contributes significantly to under-reporting. Still, parcel/plot attrition is the leading cause of the low ownership shares, while inheritance and administrative redistributions play only minor and less significant roles.

Next, we assess how well these censored Tobit models of ownership holding shares predict the actual ownership shares, given that the models were constructed to take into account real changes in ownership holding shares and parcel/plot attrition. Figure 4 presents the reported and measured versus the predicted ownership holding shares as cumulative distributions by survey round and for the panel.

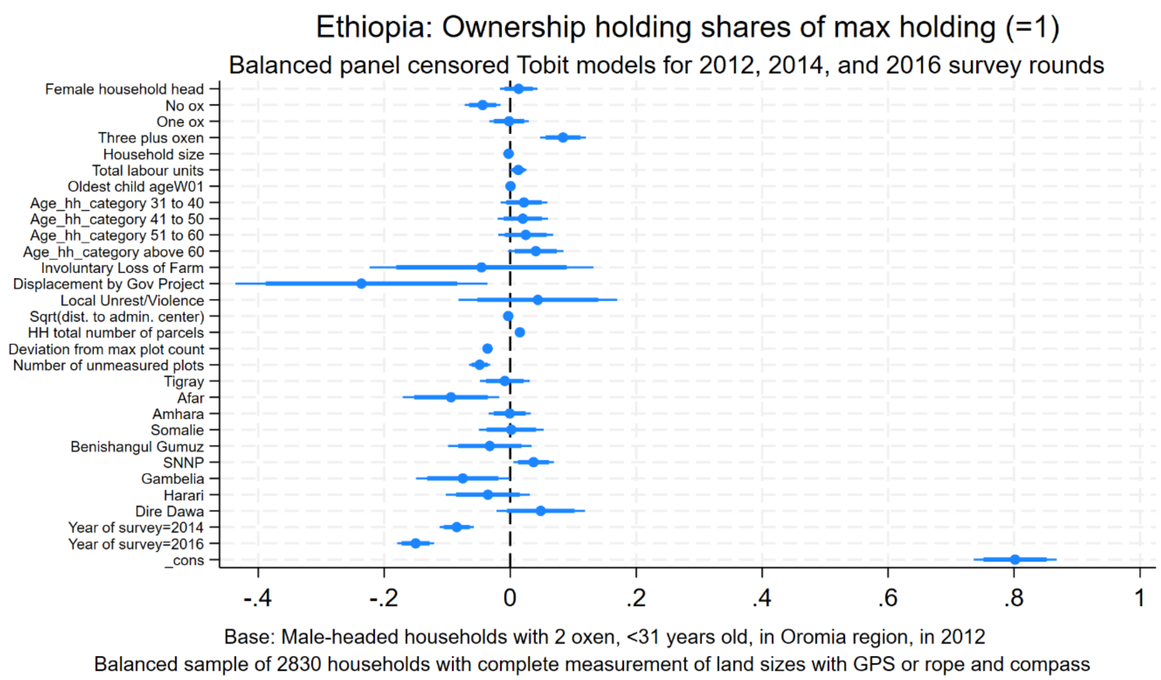

Figure 3. Panel Tobit models for ownership holding shares of maximum own ownership holding size over three survey rounds in Ethiopia

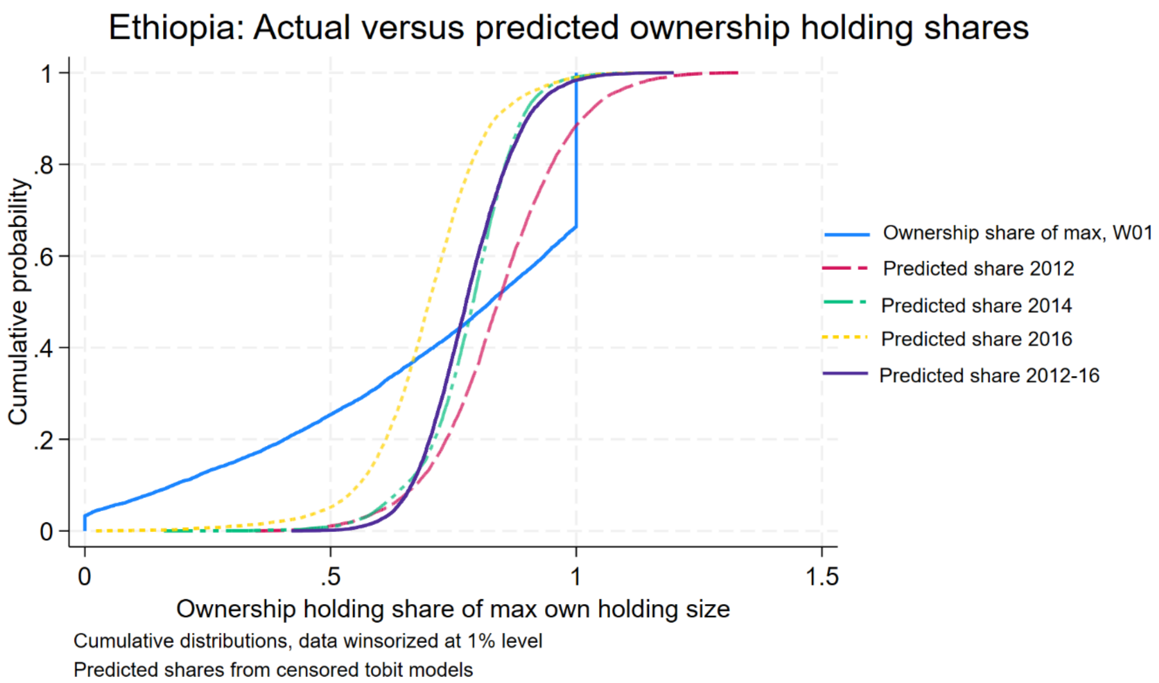

Figure 4. Ethiopia: Actual versus predicted ownership shares of max own holding sizes

The first and most fundamental problem is that the Tobit models, censored from the top, poorly predict the actual ownership holding share distribution. The model for the first 2012 survey round predicts higher ownership shares than the later survey rounds, and we found that this could partially be explained by inheritance and bequeath transfers. However, the models for 2014 and 2016 also predict poorly and provide upward-biased estimates of ownership-holding shares for the bottom one-third of the actual distribution and downward-biased estimates for most of the top half of the ownership-holding share distribution.

One may then ask whether the censored Tobit models are the problem, as they may be sensitive to non-normality and heteroskedasticity. We therefore tested alternative estimators, including fractional probit, symmetrically censored least squares estimator (SCLS), and panel stochastic frontier models. Appendix 2 presents the results of these models. We found, however, that all these estimators give poor predictions of the ownership holding shares and with similar bias distributions.

Based on this evidence, we conclude that the best proxy of households' actual holding sizes over the three survey rounds is their maximum reported and measured ownership holding size across survey rounds. This measure is the measure that is least likely to suffer from plot attrition bias. While it is not a perfect estimate, it is the best we have. We proceed by inspecting the distributions of these maximum within-household ownership holding sizes versus the actual reported and measured (with GPS or rope and compass) ownership holding sizes in the three survey rounds to get a better picture of the bias in ownership holding distributions associated with such parcel/plot attrition, see Figure 5.

Figure 5 compares the cumulative ownership holding distributions in 2012, 2014, and 2016 in ha with the cumulative within-household maximum ownership holding distribution across the three survey rounds in 2012, 2014, and 2016. Based on the previous analyses, we notice a substantial gap in all three survey rounds and suggest that this gap is primarily explained by parcel attrition that varies over time within households. The annual ownership distributions point towards >60% of farms being one ha or smaller, against <50% of the holdings being smaller than one ha according to the maximum holding distributions. We believe that the latter estimate is closer to the truth.

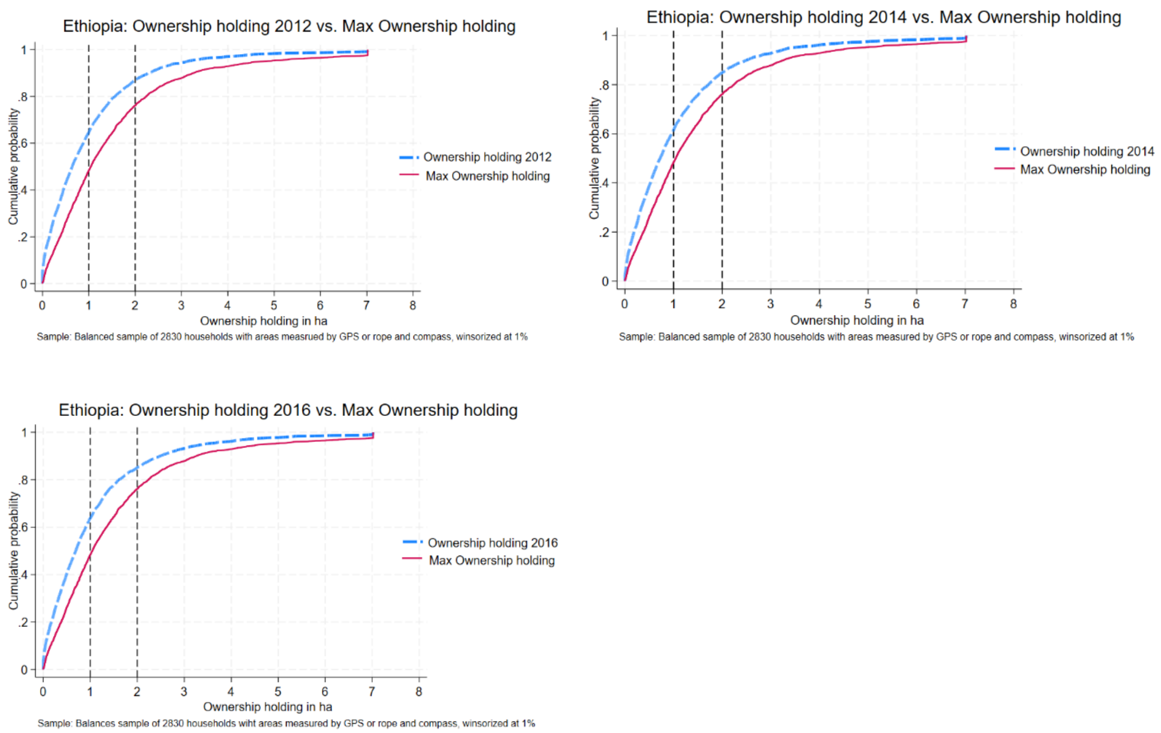

Figure 5. Ownership holding size distributions in ha in 2012, 2014, and 2016 versus maximum size distributions 2012-2016.

Next, we assess the implications for the Gini-coefficients for measured ownership holding distributions that we control for random and non-random measurement errors. Table 4 provides estimates at the regional level.

Table 4. Gini-coefficients for ownership holding size distribution by region without and with measurement error corrections

	Unwinsorized Ownership holdings 2014	1% winsorized Ownership holdings 2014	1% winsorized max ownership	2% winsorized max ownership	5% winsorized max ownership
Tigray	0.565	0.595	0.470	0.450	0.426
Afar	0.606	0.606	0.550	0.537	0.523
Amhara	0.473	0.472	0.426	0.415	0.390
Oromia	0.630	0.493	0.456	0.422	0.366
Somalie	0.763	0.494	0.513	0.500	0.466
Benishangul	0.532	0.532	0.487	0.470	0.443
SNNP	0.571	0.527	0.515	0.495	0.455
Gambella	0.656	0.651	0.551	0.551	0.532
Harari	0.444	0.438	0.426	0.400	0.369
Dire Dawa	0.387	0.387	0.378	0.367	0.349

It makes sense to adjust outlier observation values in the data as these may be driven by random measurement and aggregation errors. Table 4 illustrates that when ownership holdings are winsorized at a 1% level[iv], the Gini coefficients for the regional land distributions are substantially reduced in Oromia and Somalie regions. If we compare these estimates with the within-household maximum ownership holding winsorized at 1%, we see a further substantial reduction in the regional Gini coefficients. We suggest this is because of non-classical measurement errors due to parcel attrition. Winsorizing the maximum areas further from 1 to 2 or 5% on each side of the distributions may go too far in adjusting the tails of the maximum farm sizes (ownership holdings). Gini-coefficients at district and community levels based on land registry data in the Tigray region of Ethiopia in 2016 varied from 0.40 to 0.56 for comparison (Holden and Tilahun 2020).

4.2. Malawi

Table 5 presents the overview of the distribution of households across regions in the balanced LSMS-ISA data from Malawi. It shows that the number of households in the less densely populated Northern region is much lower than in the two more populous Central and Southern regions.

Table 5. Malawi: Distribution of the LSMS-ISA balanced household panel

Region	2010	2013	2016	2019	Total
North	86	86	86	86	344
Central	413	413	413	413	1,652
Southern	483	483	483	483	1,932
Total	982	982	982	982	3,928

Figure 6 presents the ownership holding share distributions of maximum ownership holdings for each survey round. We see a distributional pattern quite similar across survey rounds and minimal effects of winsorizing the data at 1, 2, and 5% levels. The shapes of the ownership share distributions are also astonishingly similar to those for the Ethiopian sample. This is a first indication that the parcel attrition problem we detected in the Ethiopian data also appears to be there in the Malawian data.

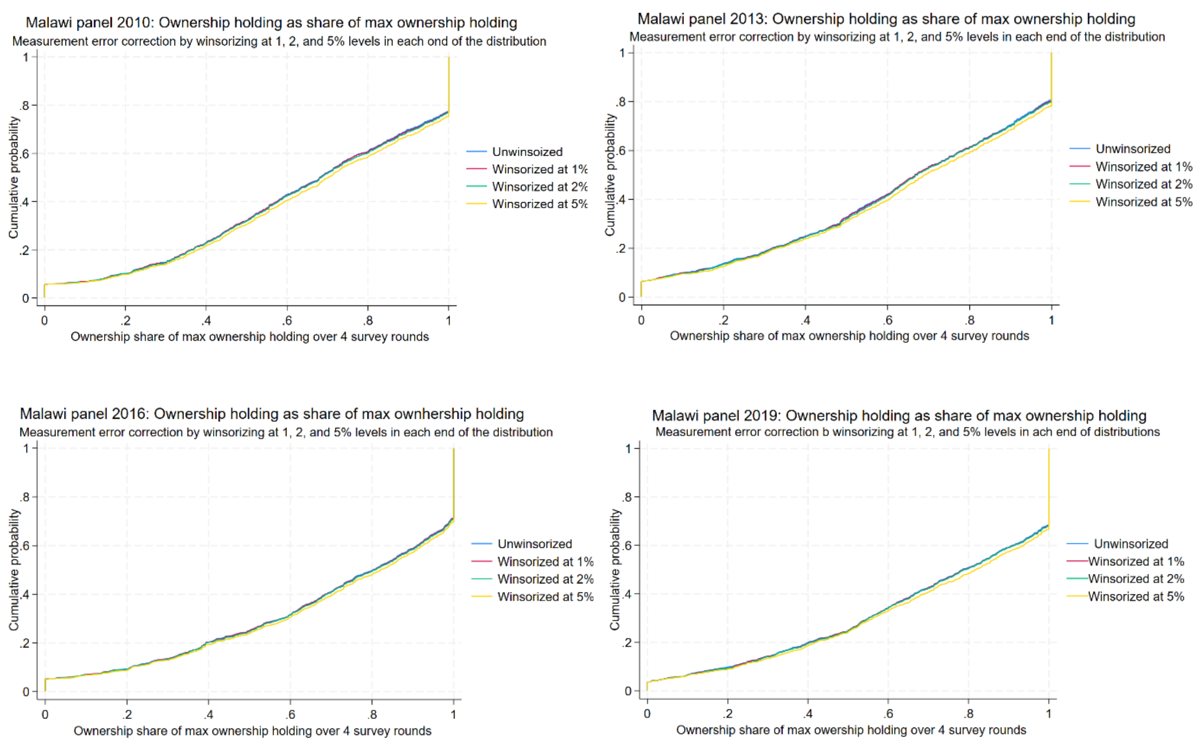

Figure 6. Within-household ownership holding shares of maximum household ownership holdings over four survey rounds in Malawi.

To scrutinize the possible parcel attrition problem, we also graph the total parcel/plot counts, and the deviation from the within-household maximum parcel/plot counts in Figures 7a and 7b. We see a deviation in parcel counts for more than 50% of the households in Figure 7a and similarly in each survey round in Figure 7b, with some indications that the problem is smaller in 2019.

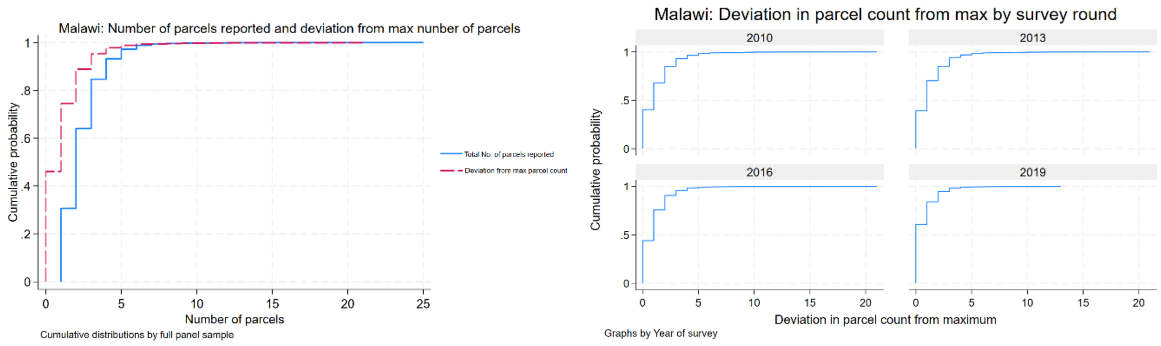

Figure 7a and 7b. Malawi: Parcel count and deviation from max parcel count (total and by survey round) for households

Based on this, we used that same estimation approach (censored Tobit models by survey round and as a four-round panel) to detect factors associated with real changes in within-household ownership holding shares over time and factors associated with varying degrees of parcel/plot attrition. The results for all models are found in Table 6.

Table 6 results indicate that splitting in age classes for household heads did not provide strong evidence that bequeath and inheritance can explain much of the low ownership holding shares. However, there is evidence that the youngest age classes have higher holdings in the last (2019) survey round. The constant terms in the Malawi models are lower than in the Ethiopian models and illustrate a sizeable average gap for the baseline household category. The variables, which are potential indicators of land being rented out (female head, total labor force, drought power dummy), were only significant in the 2019 model and the full panel. On the other hand, the variables potentially indicating parcel/plot attrition (total plot number reported, deviation from maximum parcel/plot count) causing low ownership holding shares were highly significant and with positive and negative signs, in line with these variables signaling significant under-reporting of parcels/plots for a large share of the households compared to the year they reported the most complete parcel/plot counts (Figure 7b).

Table 6. Malawi: Tobit models for ownership holding shares of max holding shares by year

	t1	t2	t3	t4	t14
	b/se	b/se	b/se	b/se	b/se
Year→	2010	2013	2016	2019	2010-2019
Female headed, dummy	0.010	-0.007	-0.052	0.051	0.004
	(0.04)	(0.04)	(0.04)	(0.04)	(0.01)
Total labor units	-0.021	0.009	0.003	0.048****	0.011**
	(0.02)	(0.02)	(0.02)	(0.01)	(0.01)
Livestock endowment	-0.000	0.004	-0.001	-0.005***	-0.001
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Draught_power, dummy	0.078	-0.003	0.005	0.127*	0.052**
	(0.08)	(0.06)	(0.06)	(0.07)	(0.03)
Oldest child age	0.000	-0.002	-0.001	-0.005***	-0.001*
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Age hhh <31 (Base)					
Age hhh 31-40	-0.007	0.038	0.112	0.166***	0.056***
	(0.05)	(0.05)	(0.07)	(0.06)	(0.02)
Age hhh 41-50	0.004	0.023	0.030	0.145**	0.026
	(0.06)	(0.05)	(0.08)	(0.06)	(0.02)
Age hhh 51-60	0.055	-0.021	0.122	0.062	0.029
	(0.05)	(0.07)	(0.10)	(0.07)	(0.02)
Age hhh >60	0.105*	0.087	0.114	0.114	0.077****
	(0.06)	(0.06)	(0.08)	(0.08)	(0.02)
Distance pop.center	-0.001	-0.001	0.000	0.001	0.000
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Parcel count	0.072****	0.064****	0.074****	0.079****	0.075****
	(0.02)	(0.02)	(0.01)	(0.01)	(0.01)
Deviation max parcel count	-0.050****	-0.038****	-0.085****	-0.054****	-0.057****
	(0.01)	(0.01)	(0.02)	(0.01)	(0.00)
Parcels not measured, dummy		-0.459****	-0.085	-0.324****	-0.278****
		(0.07)	(0.17)	(0.09)	(0.05)
Urban_rural, dummy	0.026	0.053	0.020	0.085	0.049*
	(0.08)	(0.06)	(0.07)	(0.06)	(0.03)
Northern region (base)	0.000	0.000	0.000	0.000	0.000
Central region	-0.057	-0.03	0.123***	0.012	0.024
	(0.08)	(0.05)	(0.04)	(0.05)	(0.02)
Southern region	-0.034	0.047	0.144***	0.061	0.071***
	(0.07)	(0.05)	(0.04)	(0.05)	(0.02)
2010.panel year(base)					0.000
2013.panel year					-0.043***
					(0.01)
2016.panel year					0.057****
					(0.010)
2019.panel year					-0.027*
					(0.02)
Constant	0.664****	0.400****	0.379***	0.303**	0.454****
	(0.12)	(0.10)	(0.15)	(0.15)	(0.05)

var(e.own~1)	0.127****	0.118****	0.141****	0.128****	
	(0.01)	(0.01)	(0.01)	(0.01)	
sigma_u					0.000
					(0.01)
sigma_e					0.364****
					(0.00)
Prob > chi2	0.0000	0.0000	0.0000	0.0000	0.0000
Number of observations	982	982	982	982	3928

*Note:*Models corrected for attrition with inverse probability weighting (IPW). Standard errors in parentheses. Significance levels: *: 10% level, **: 5% level, ***: 1% level, ****: 0.1% level.

A one-unit deviation in plot count from the maximum within-household plot count is associated with between 3.8 and 8.5 percentage point smaller ownership holding shares across the sample round models, with all these estimates being highly significant (0.1% levels). We also included a dummy for households with unmeasured parcels/plots, which were not included in the aggregate household ownership holding measure. As expected, this dummy was associated with a large and significant (at 1% level) negative effect on the ownership holding share. However, very few households reported such unmeasured parcels/plots, thus this variable can only explain small ownership holding shares for a very small share of the sample. The results in Table 6 combined with Figure 7b indicate that plot attrition is the main driver of the low ownership holding shares.

Figure 8 presents the results for the censored panel Tobit model, which confirm that variation in parcel/plot reporting is the main reason for the variation in ownership holdings. The variations over survey rounds were small, indicating that the problem persists. We predicted the ownership holding shares based on the censored Tobit models for each survey round and on the panel Tobit model. We compare the cumulate probability distributions for the predicted ownership holding shares with the actual ownership shares in Figure 9. The figure shows the same poorly predicted fits to the actual ownership holding share distributions as we saw for the Ethiopian data. We come to the same conclusion as for the Ethiopian data that the maximum within-household farm size over survey rounds is a better proxy for the average ownership holding size than the measures for each survey round, as this maximum holding size is the least likely to suffer from parcel attrition.

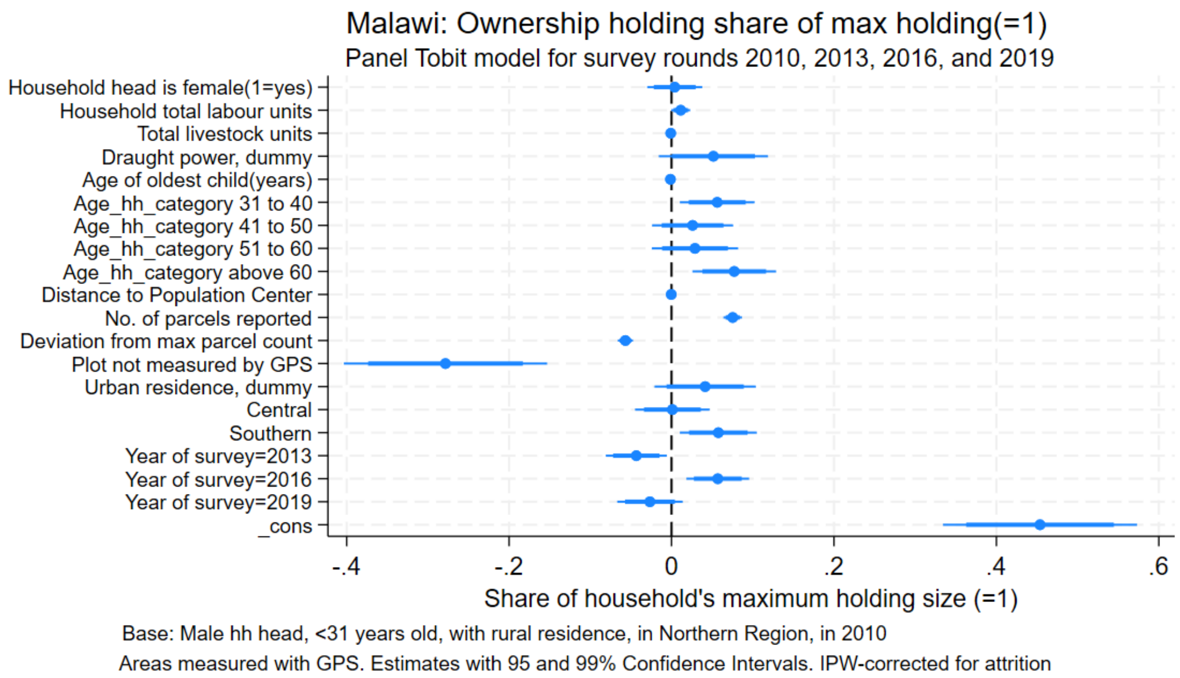

Figure 8. Malawi: Panel Tobit model estimates for ownership holding shares in 2010-2019 panel

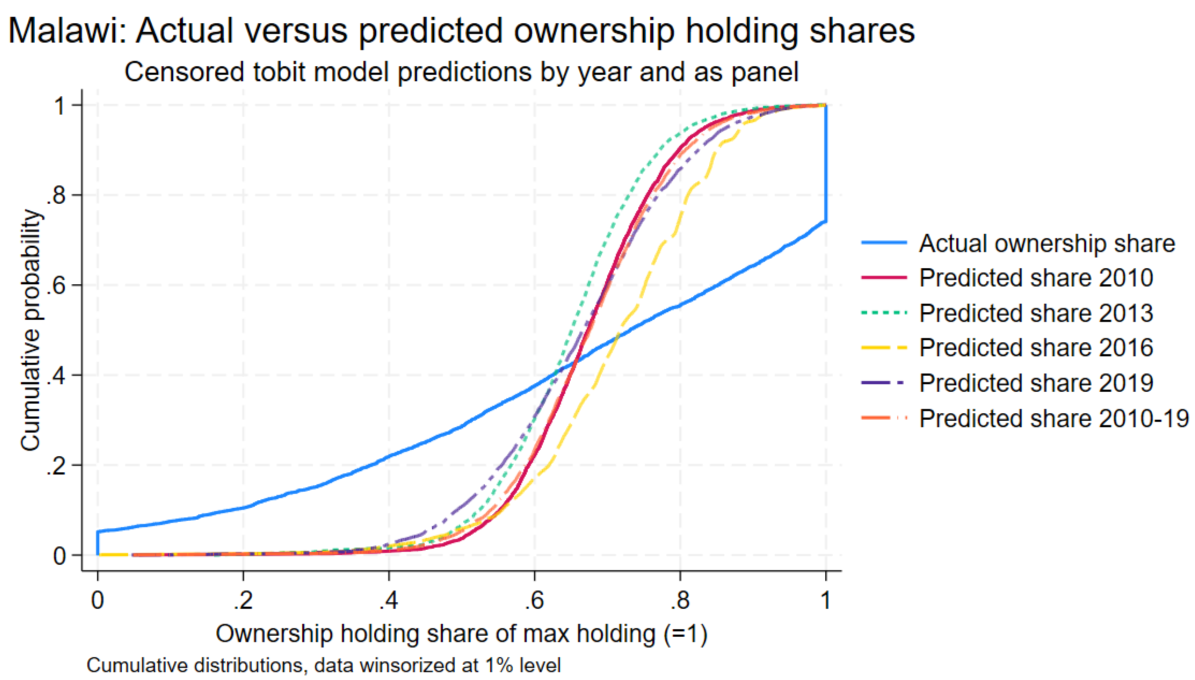

Figure 9. Malawi: Actual versus predicted owner ownership shares from Tobit models

Based on this assessment, we proceed as we did with the Ethiopian data by comparing the maximum holding sizes measured in ha with the actual holding sizes in each survey round. We complement the analyses by alternatively winsorizing the reported and measured farm sizes at 1% level at each distribution end. We use 1, 2, and 5% winsorizing for the maximum holding sizes, see Table 7.

Table 7 shows that the annual per survey round holding sizes are much smaller than the maximum holding sizes, whether winsorized or not. The reported and measured ownership holding size distributions in each survey round and the maximum holding size distribution, winsorized at 1% level, are presented in Figure 10. We claim that these graphs give a good picture of the downward bias in farm sizes caused by the stochastic attrition in the reporting of parcels/plots in these surveys in the case of Malawi.

Figure 11 illustrates the regional variation in these reported and measured farm size distributions versus the maximum ownership holding size distribution over the panel years. We see that the gap is largest in the Northern region where the sample is the smallest and the population densities are the lowest (making fallowing more likely as an additional reason for under-reporting of parcels/plots).

Table 7. Estimated ownership holding sizes in ha based on 4 rounds of household-parcel panel data from Malawi, without and with outlier corrections.

	Ownership holding, unwinsorized				Ownership holding, winsorized 1%				Max ownership holding, winsorized at		
Year	2010	2013	2016	2019	2010	2013	2016	2019	1%	2%	5%
Mean	0.738	0.697	0.803	0.788	0.713	0.690	0.793	0.781	1.133	1.102	1.025
Median	0.567	0.532	0.631	0.599	0.567	0.532	0.631	0.599	0.902	0.902	0.902
P25	0.308	0.291	0.336	0.332	0.308	0.291	0.336	0.332	0.587	0.587	0.587
P75	0.902	0.890	1.036	1.036	0.902	0.890	1.036	1.036	1.449	1.449	1.449
P90	1.425	1.392	1.639	1.619	1.425	1.392	1.639	1.619	2.125	2.125	1.947
St.dev.	0.837	0.652	0.741	0.718	0.626	0.609	0.688	0.680	0.790	0.702	0.543
St.err.	0.027	0.021	0.024	0.023	0.020	0.019	0.022	0.022	0.025	0.022	0.017
N	986	986	986	986	986	986	986	986	986	986	986
Gini	0.416	0.407	0.415	0.420	0.396	0.401	0.407	0.415	0.358	0.341	0.298

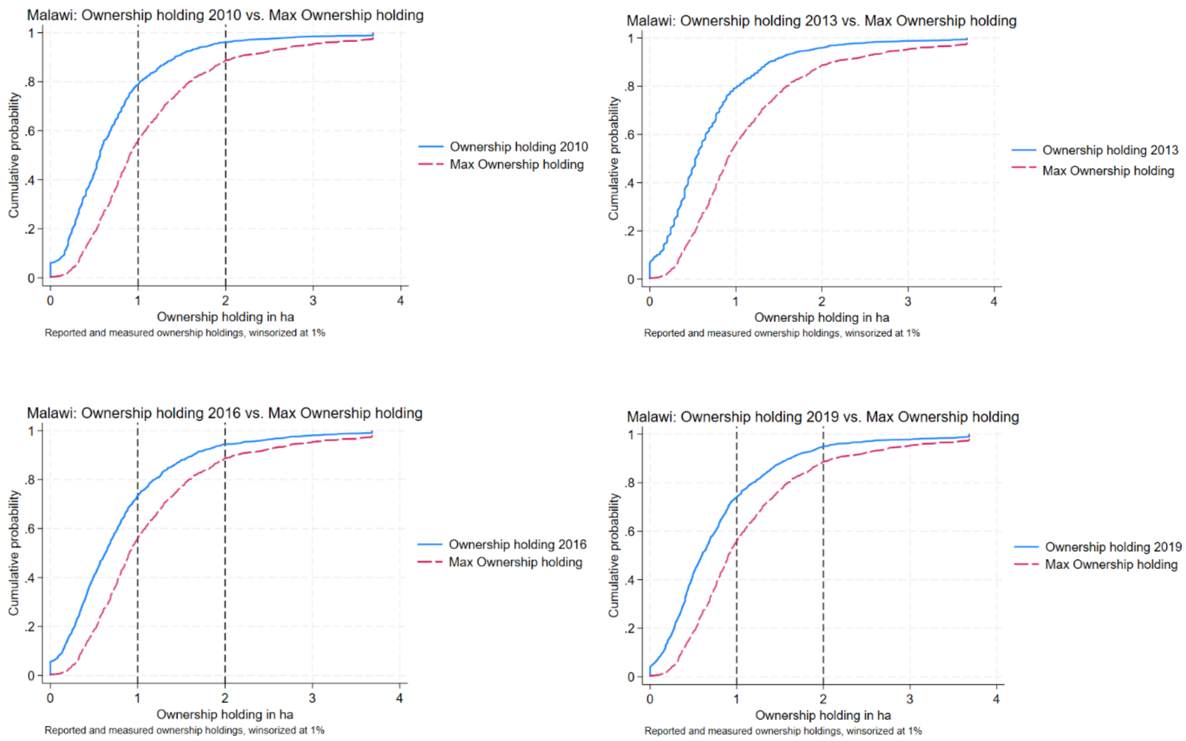

Figure 10. Ownership holding distributions by survey year versus maximum within-household ownership holdings across the four survey rounds.

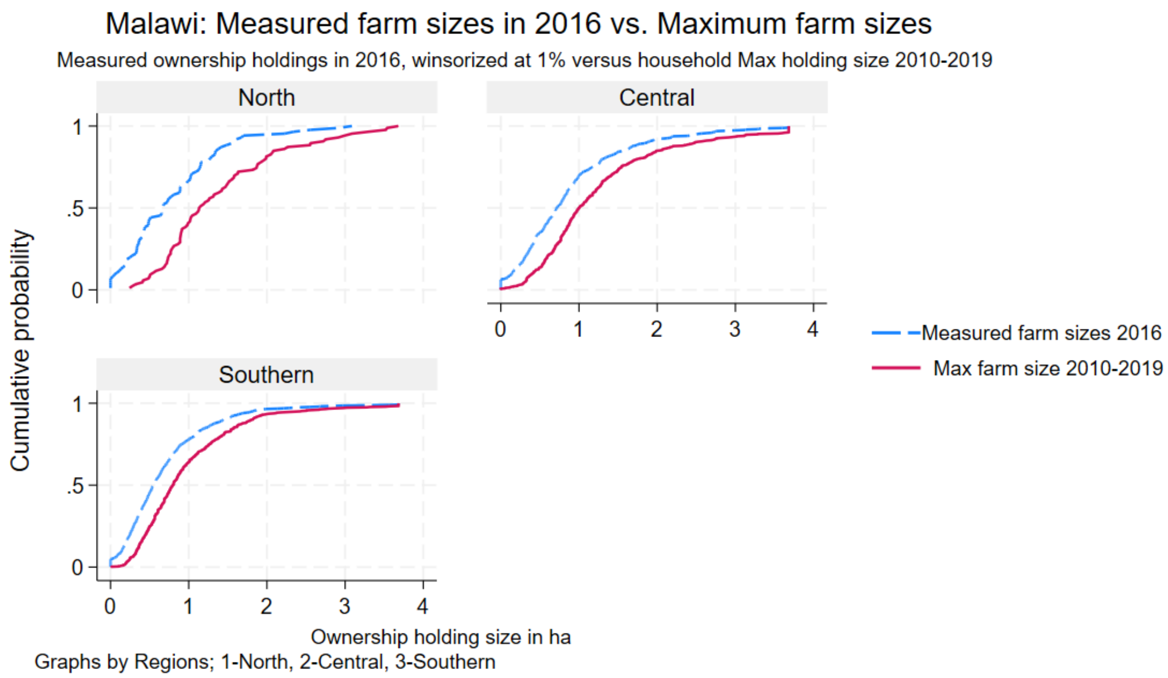

Figure 11. Measured farm sizes in 2016 versus max farm size 2010-19 by region

Finally, we assess the Gini distributions for the actual and maximum ownership holding sizes in Table 8. The table indicates that the parcel/plot attrition leads to an upward bias in the estimated Gini coefficients for farm size distributions, like we found in Ethiopia.

Table 8. Malawi: Gini-coefficients by region, without and with winsorized variables

Region	Ownership holdings Unwinsorized	Ownership holdings Winsorized at 1%	Max Ownership holding Winsorized at 1%	Max Ownership holding Winsorized at 2%	Max Ownership holding Winsorized at 5%
North	0.400	0.399	0.313	0.297	0.242
Central	0.409	0.395	0.351	0.330	0.278
Southern	0.413	0.406	0.358	0.345	0.314
Total	0.416	0.406	0.358	0.341	0.298

5. Discussion

It is very demanding to collect reliable farm size data in field surveys, whether measured with handheld GPSs or rope and compass. Such data collection is unnecessary in countries where reliable land registries are linked to digital maps. While such a digital and administrative revolution is underway in an increasing number of developing countries, reliable agricultural statistics are vital in countries without such land registry data. We have investigated the reliability of the measured farm sizes in the nationally representative LSMS-ISA surveys in Ethiopia and Malawi and have found substantial downward bias in the estimated farm sizes, caused mainly by systematic and stochastic parcel/plot attrition. We were able to detect this type of error by combining multiple survey rounds from the same balanced sample of farm households in the two countries. Earlier studies aiming to enhance the reliability of the measurement of farm and parcel sizes have utilized data from single survey rounds and compared alternative techniques of measuring parcel sizes. However, the phenomenon we studied was not detectable with that approach.

We have benefitted from access to data from three survey rounds in Ethiopia and four in Malawi. We find that the extent of the problem is substantial in both countries. We provided some theoretical propositions for why we thought this might be an essential problem, and we believe that those propositions may explain the problem and the large biases in farm size estimates that we have detected.

Our findings have important implications for policy analyses based on these data and for developing better ways to generate reliable measures of farm sizes in these countries.

For policy analysis purposes, it implies that these surveys do not provide reliable measures of ownership holding sizes in each survey round, even if all reported parcels have been measured with GPS. The under-reporting of parcels/plots leads to the under-reporting of farm sizes. Further work is needed to assess how this may affect the reporting of outputs from and input use on unreported parcels. If such data are also collected at the plot/parcel level, plot/parcel attrition will also lead to a similar downward bias in reported output and input data at the household level. The bias may be less if input and output data are collected at the household level or have been through quality and consistency checks across plot, parcel, and household levels. This is an important area for future follow-up research.

The benefit of the balanced household panels is that we may use the maximum holding sizes over survey rounds to get more reliable measures of actual farm sizes. One may also combine these maximum holding sizes with real identified changes in farm sizes associated with inheritances and bequeaths, purchases and sales of land, and administrative redistributions to get more exact farm size distributions. Such refinements go beyond this paper's scope, which was to assess the importance of this type of measurement error. It is also evident that with this type of measurement error, it becomes even more tricky to study the famous relationship between farm size and land productivity as the under-reporting of areas and possibly the related input use and outputs for unreported parcels cannot be assumed to represent white noise in the data.

Researchers within the World Bank or the Statistical Offices within Ethiopia and Malawi have access to the detailed GPS coordinates at the parcel level. These data could be used to dig deeper and create parcel panel data to study further the parcel/plot attrition and how parcels have, in different ways, been split into different sub-parcel/plot structures over the years within households. Based on the household-level Parcel IDs across survey rounds, we inspected how well parcel sizes matched across years within households for selected households along the ownership holding share distributions (Appendix 4). These findings indicate that the Parcel IDs do not match well across years within households, even though the enumerator instructions had emphasized that such matching should be implemented.

Another obvious way forward could be to explicitly link the survey households to the land registry data likely to exist for a large share of the households in the Ethiopian sample. Ethiopia has undergone two land registration processes during the last 10-27 years, and a large share of the rural households, therefore, have land certificates for their ownership holdings. The second-stage land certification process, which took place around the time of the Ethiopian surveys (2014-16), should provide reliable farm size data for a large share of the sample households. Linking the LSMS-ISA data to the land registry data would require access to household IDs and the land registry data and could be an interesting control of the findings in our study.

In smaller surveys in various locations in Ethiopia, the first author trained the enumerators to ask to see the land certificates of the households and to record the parcel and farm-level data from these certificates. Unfortunately, the LSMS-ISA surveys did not do the same in Ethiopia.

Another way to get more reliable parcel and farm size data is to use a more comprehensive land tenure module as part of the LSMS-ISA survey instrument (Holden et al. 2016). While this has been proposed, it has not been implemented to any extent due to the already huge survey instrument used in these multipurpose surveys. The survey instrument represents a heavy burden for the responding households and trained enumerators. Unsurprisingly, they have incentives to under-report in collecting parcel/plot level data that also require visits to the field for parcel/plot measurement. An easier way to enhance the reliability of the parcel-level data would have been to have the tablets preprogrammed with the parcel data from the previous survey round. A potential danger could be that an initial parcel attrition type of error could carry over from the first round, and the balanced panel could become less suitable for detecting this type of error.

Protecting respondents' anonymity is a crucial issue to handle carefully related to such balanced panel household surveys. However, a core group of persons must manage the household, person, and GPS coordinate data. They should also receive training to deal more effectively with these measurement errors to help generate more reliable agricultural statistics.

6. Conclusions

We have used three LSMS-ISA survey rounds for Ethiopia and four survey rounds for Malawi to construct balanced household-farm data to assess the reliability of the estimated farm ownership holding sizes of these households. Almost all the recorded parcels and sub-parcels/plots have been measured with reliable tools such as handheld GPSs and/or rope and compass, giving reliable estimates of the recorded areas. Our contribution to the literature is to use the balanced panel to investigate the within-household variation in reported parcels and, thereby, ownership holding sizes over several survey rounds and investigate the possible reasons for such a variation over time. This variation could be either real and caused by inheritances and bequeaths, purchases and sales, administrative redistributions, or private land grabs. However, we found that much of this variation was associated with varying degrees of under-reporting parcels/plots over time. One reason for under-reporting was a lower likelihood of reporting land rented out and thereby being farmed by somebody else. It could also be due to other reasons for convenience for hiding areas related to the drudgery of reporting all relevant data associated with the reported parcels, including going to the field and measuring the parcels/plots.

We found this under-reporting of parcels/plots caused large non-classical and econometrically difficult-to-predict measurement errors. While the econometric models revealed the severity of the problem, they could not adequately correct for the severe detected attrition bias associated with under-reported parcels/plots. We conclude that the best and easiest proxy variable for the actual farm size of households is the maximum within-household reported and measured ownership holding size over the survey rounds. This maximum holding could also be biased downward due to under-reporting but much less so than the GPS-measured ownership holding sizes in each survey round. We demonstrate the degree of bias by comparing actual reported and measured holding size distributions versus the distribution of the maximum within-household holdings. These discrepancies are substantial in all survey rounds and across all regions in both Ethiopia and Malawi. The ignorance of the biases due to such parcel/plot attrition causes average and median ownership holdings to be underestimated by 12-41% and Gini coefficients for ownership holding distributions to be substantially over-estimated.

Here are important policy implications regarding the need to take these substantial non-classical measurement errors into account when using these data for policy analyses to generate aggregate data at the national level. Likewise, the statistical authorities responsible for these surveys and future data collection must consider this measurement error problem. Our clear perspective is that these errors have largely gone under the radar as they are easy to overlook when focusing on the data from a single survey round. We have provided recommendations for alternative ways to reduce the problem in future surveys and further scrutinize the errors we detected in the collected data. The similarity in the findings in these two LSMS-ISA countries makes us reasonably confident that this problem also exists in other LSMS-ISA and similar surveys in different countries. Investigating this should be an area for future research to generate more reliable agricultural statistics for improving and developing future agrarian policies.

References

Abay, K.A., Abate, G.T., Barrett, C.B. and Bernard, T., (2019). Correlated non-classical measurement errors, 'Second best' policy inference, and the inverse size-productivity relationship in agriculture. *Journal of Development Economics*, 139, 171-184.

Abay, K.A., Bevis, L.M., Barrett, C.B., (2021). Measurement error mechanisms matter: agricultural intensification with farmer misperceptions and misreporting. *American Journal of Agricultural Economics*, 103, 498–522.

Abay, K.A., Wossen, T., Abate, G.A., Stevenson, J., Michelson, H., Barrett, C.B., (2023). Inferential and behavioral implications of measurement error in agricultural data. *Annual Review of Resource Economics*, 15, 1.

Abay, K.A., Barrett, C.B., Kilic, T., Moylan, H., Ilukor, J. and Vundru, W.D., 2023. Non-classical measurement error and farmers' response to information treatment. *Journal of Development Economics*, p.103136.

Carletto, C. (2021). Better data, higher impact: improving agricultural data systems for societal change. *European Review of Agricultural Economics*, 48 (4), 719–740

Carletto, C., Gourlay, S., & Winters, P. (2015). From guesstimates to GPStimates: Land area measurement and implications for agricultural analysis. *Journal of African Economies*. First published online May 29, 2015. doi:10.1093/jae/ejv011

Carletto, C., Gourlay, S., Murray, S. and Zezza, A., (2017). Cheaper, faster, and more than good enough: Is GPS the new gold standard in land area measurement? *Survey Research Methods* 11(3), 235-265.

Carletto, C., Savastano, S., & Zezza, A. (2013). Fact or artifact: The impact of measurement errors on the farm size–productivity relationship. *Journal of Development Economics*, 103(100), 254–261.

Carletto, C., Dillon, A. and Zezza, A. (2021). Agricultural data collection to minimize measurement error and maximize coverage. In *Handbook of Agricultural Economics*, 5, 4407-4480. Elsevier.

Burke, W.J., Morgan, S.N., Namonje, T., Muyanga, M. and Mason, N.M. (2019). *Area Mismeasurement Impact on Farmers' Input Choices and Productivity* (No. 1878-2020-525). Michigan State University, Department of Agricultural, Food, and Resource Economics, Feed the Future Innovation Lab for Food Security (FSP).

Gourlay, S., Kilic, T. and Lobell, D.B. (2019). A new spin on an old debate: Errors in farmer-reported production and their implications for inverse scale-productivity relationship in Uganda. *Journal of Development Economics*, 141, p.102376.

Holden, S. T., Otsuka, K. and Place, F. (Eds) (2008). The Emergence of Land Markets in Africa: Impacts on Poverty, Equity and Efficiency. Resources For the Future Press, Washington D.C.

Holden, S. T., Deininger, K. and Ghebru, H. (2011). Tenure Insecurity, Gender, Low-cost Land Certification and Land Rental Market Participation. *Journal of Development Studies* 47(1), 31-47.

Holden, S.T., Ali, D., Deininger, K. and Hilhorst, T. (2016). A Land Tenure Module for LSMS. CLTS Working Paper No. 1/2016. Center for Land Tenure Studies, Norwegian University of Life Sciences, Aas, Norway.

Holden, S.T., Fisher, M. (2013). Can Area Measurement Error Explain the Inverse Farm Size Productivity Relationship? CLTS Working Paper No. 12 Norwegian University of Life Sciences, Ås, Norway.

Holden, S.T., and Tilahun, M. (2020). Farm size and gender distribution of land: Evidence from Ethiopian land registry data. *World Development*, 130, p.104926.

Jayne, T.S., Muyanga, M., Wineman, A., Ghebru, H., Stevens, C., Stickler, M., Chapoto, A., Anseeuw, W., van der Westhuizen, D. and Nyange, D. (2019). Are medium-scale Farms Driving Agricultural Transformation in Sub-Saharan Africa? *Agricultural Economics*, 50, 75-95.

Kilic, T., Yacoubou Djima, I. and Carletto, C. (2017). Mission impossible? Exploring the promise of multiple imputation for predicting missing GPS-based land area measures in household surveys. *Exploring the Promise of Multiple Imputation for Predicting Missing Gps-Based Land Area Measures in Household Surveys (July 6, 2017). World Bank Policy Research Working Paper*, (8138).

Kilic, T., Zezza, A., Carletto, G., & Savastano, S. (2016). Missing(ness) in Action: Selectivity Bias in GPS-Based Land Area Measurements. *World Development, 92*. doi:10.1016/j.worlddev.2016.11.018

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, 16‑29.

World Bank. (2021). World development report: data for better lives. Washington, DC: World Bank. https://www.worldbank.org/en/publication/wdr2021. Accessed 29 March 2021.

Wossen, T., Abay, K.A. and Abdoulaye, T., 2022. Misperceiving and misreporting input quality: Implications for input use and productivity. *Journal of Development Economics*, 157, p.102869.

[i] Both these surveys are part of the Livings Standards Measurement Study – Integrated Surveys on Agriculture (LSMS-ISA) program.

[ii] These surveys are population-based and may therefore not adequately capture medium- and large-sized farms (Jayne et al. 2019).

[iii] We use the terms sub-parcel and plot interchangeably. Parcels have more stable borders in the landscape while sub-parcel or plot borders may vary from year to year depending on cropping pattern.

[iv] Winsorizing was done for the whole sample and not region by region.

Missing Parcels and Farm Size Measurement Error: -Do Nationally Representative Surveys Provide Reliable Estimates?

Appendices/Supplementary materials

Appendix 1. Descriptive statistics by country

Table A1.1 below presents the basic descriptive statistics for the variables included in the econometric models for the Ethiopian balanced panel. Six households reported no owned land in all three survey rounds and are, therefore, dropped in the ownership holding share models.

Table A1.1. Right-hand side variables used in the Ethiopian panel data models, by survey round

	2012		2014		2016	
Variable	Mean	St.Err.	Mean	St.Err.	Mean	St.Err.
Female-headed hh	0.183	0.008	0.199	0.008	0.210	0.008
No ox	0.445	0.010	0.460	0.010	0.457	0.010
One ox	0.224	0.009	0.212	0.008	0.212	0.008
Two oxen	0.260	0.009	0.268	0.009	0.256	0.009
Threeplus oxen	0.070	0.005	0.061	0.005	0.075	0.005
Household size	5.345	0.045	6.007	0.048	6.504	0.050
Total labor units	3.874	0.032	3.608	0.031	3.893	0.036
Oldest child age	15.913	0.127	15.903	0.160	16.229	0.150
Age hhh 20-30	0.199	0.008	0.147	0.007	0.103	0.006
Age hhh 31-40	0.258	0.009	0.259	0.009	0.248	0.009
Age hhh 41-50	0.221	0.009	0.233	0.009	0.252	0.009
Age hhh 51-60	0.150	0.007	0.170	0.008	0.176	0.008
Age hhh > 60	0.171	0.008	0.191	0.008	0.221	0.009
Involunatry loss of farm, dummy	0.004	0.001	0.003	0.001	0.004	0.001
Displacement by Gov., dummy	0.002	0.001	0.001	0.001	0.003	0.001
Local unrest shock, dummy	0.003	0.001	0.001	0.001	0.017	0.003
Sqrt(Distance to admin center)	11.866	0.099	11.900	0.098	11.896	0.098
Parcel/plot count	3.787	0.057	4.426	0.067	4.941	0.073
Deviation from max plot count	1.252	0.037	0.614	0.023	0.098	0.010
Number of unmeasured parcels	0.075	0.011	0.467	0.023	0.054	0.007

Table A1.2 shows descriptive statistics aggregated to the parcel level for the balanced sample by year. As indicators of parcel attrition associated with renting, we see that renting in is much more frequently reported than renting out. Sharecropping was only explicitly recorded in the

2016 round.

Table A1.2. Ethiopia: Descriptive statistics for selected variables at parcel-level

Variable definitions	2012		2014		2016	
	mean	sd	mean	sd	mean	sd
Parcels acquired from local leader(1=yes)	0.40	0.49	0.40	0.49	0.38	0.49
Parcel acquired through inheritance(1=yes)	0.36	0.48	0.44	0.50	0.44	0.50
Parcel acquired through renting(1=yes)	0.13	0.34	0.13	0.33	0.06	0.25
Parcel acquired through borrowing(1=yes)	0.05	0.21	0.03	0.16	0.02	0.14
Parcel acquired through moving in without permission(1=yes)	0.05	0.22	0.04	0.20	0.04	0.19
Parcel acquired through sharecropping (IN)(1=yes)	0.00	0.00	0.00	0.00	0.08	0.27
Parcel acquired through purchase(1=yes)	0.00	0.00	0.00	0.00	0.04	0.21
Parcel acquired through other means(not-specified) (1=yes)	0.06	0.23	0.05	0.21	0.01	0.12
Parcel has a certificate(1=yes)	0.36	0.48	0.44	0.50	0.49	0.50
Parcel has some fields rented out? (1=yes)	0.01	0.11	0.05	0.21	0.03	0.16
Parcel is new in this round(1=yes)	0.00	0.00	0.18	0.38	0.13	0.34
Parcel measured by GPS(1=yes)	0.88	0.33	0.90	0.30	1.00	0.03
Parcel measured by rope or compass(1=yes)	0.24	0.43	0.27	0.44	0.14	0.35
Parcel measured by GPS and or Rope/Compass(1==yes)	0.94	0.24	0.91	0.29	1.00	0.02
Total number of sub-Parcels(fields) recorded(counted)	3.11	4.00	2.75	3.29	2.82	3.43
Total number of sub-Parcels(fields) self-reported	3.34	4.24	2.73	3.32	2.84	3.56
Total number of sub-parcels (fields) measured by GPS	2.72	4.00	2.62	3.36	2.82	3.43
Total number of sub-parcels (fields) measured by GPS and or ROPE/COMPASS	3.02	4.03	2.63	3.35	2.82	3.43
Total number of sub-parcels (fields) not measured by GPS/Compass	0.11	0.70	0.12	0.50	0.00	0.09
Mean distance from household to Parcel(km)	2.48	33.60	1.52	6.33	1.11	8.01
Mean distance from household to Parcel(km) outliers curbed at 2%	0.84	1.20	0.91	1.38	0.90	1.28
Households	2846		2846		2846	
Parcels	9810		10642		10320	

Table A1.3 presents the detailed descriptive statistics at the household level for the Malawi panel by survey round. The table shows, among others, that only a tiny share of the reported plots were not measured by GPS.

Table A1.3: Descriptive statistics at the household level Malawi sample 2010-2019

Variable definitions	2010		2013		2016		2019	
	mean	sd	mean	sd	mean	sd	mean	sd
Total No. of plot reported by household	2.120	1.088	2.149	1.108	1.380	0.797	1.578	0.929
Operational holding (GPS) in ha	0.725	0.838	0.807	1.155	0.774	0.721	0.772	0.711
Operational holding (self-reported) in ha	0.759	0.753	0.748	0.671	0.774	0.661	0.759	0.621
Ownership holding (GPS) in ha	0.654	0.801	0.745	1.162	0.725	0.751	0.726	0.727
Ownership holding (self-reported) in ha	0.694	0.725	0.686	0.662	0.728	0.677	0.709	0.647
Plot not measured by GPS(1=yes)	0.000	0.000	0.019	0.138	0.013	0.114	0.021	0.144
Reason: plot was too far(1=yes)	0.000	0.000	0.009	0.095	0.999	0.032	0.995	0.071
Reason: household refused(1=yes)	.	.	0.005	0.071	0.002	0.045	0.002	0.045
Reason: plot was waterlogged(1=yes)	0.000	0.000	.	.	.	.	.	.
Reason: other (1=yes)	.	.	0.006	0.078	0.004	0.064	0.019	0.138
Household size	5.032	2.226	5.548	2.304	5.463	2.304	5.251	2.319
Age of household head (years)	43.768	16.022	46.239	15.716	48.458	15.297	50.130	14.604
below 30 years	0.247	0.432	0.168	0.374	0.096	0.295	0.051	0.220
31-40 years	0.251	0.434	0.273	0.446	0.262	0.440	0.235	0.424
41 to 50 years	0.193	0.395	0.206	0.405	0.231	0.422	0.298	0.458
51 to 60 years	0.138	0.345	0.153	0.360	0.186	0.389	0.166	0.373
Above 60 years	0.171	0.377	0.200	0.400	0.225	0.418	0.249	0.433
Household head if female (1=yes)	0.244	0.430	0.260	0.439	0.312	0.464	0.299	0.458
Household total labour units (male adult equivalent)	2.876	1.312	3.247	1.366	3.390	1.442	3.288	1.431
Total livestock units	0.453	3.415	0.417	2.374	0.486	3.950	0.463	2.935
Household has animals which can be used as draught power (oxen, donkey, mule, etc(1=yes))	0.035	0.185	0.047	0.211	0.053	0.224	0.053	0.224
Number of draught animals the household has	0.266	4.152	0.260	2.790	0.373	5.319	0.327	3.448
Distance to ADMARC (km)	7.389	5.090	7.567	5.207	7.595	5.226	7.373	5.445
Distance to a paved road(km)	9.689	9.985	9.715	10.110	9.889	10.117	9.176	9.649
Average distance from household to plot(km)	0.873	1.207	1.704	4.972	1.845	5.826	1.878	5.629
Age of oldest child(years)	15.296	11.659	15.510	8.103	16.688	7.407	17.510	8.595

Appendix 2. Robustness checks with alternative estimators

To inspect whether alternative estimators to the censored Tobit models give different results, we tested fractional probit and symmetrically censored least squares (SCLS) estimators for Ethiopia's 2016 survey round data. Figure A2.1 shows some variation across the explanatory variables in terms of the size and 95 and 99% confidence intervals for the included variables. The fractional probit had clearly larger confidence intervals than the other models. Figure A2.2 shows the predicted ownership holding shares for the three models. The predicted distributions by SCLS and fractional probit models were very close to each other. Still, they demonstrated a poorer fit than the censored Tobit model for the upper 60% of the distribution of actual ownership holding shares. However, none of the estimators are good at predicting the very high and very low ownership holding shares.

Next, we have compared the performance of panel censored Tobit models and panel stochastic frontier models for the fully balanced Ethiopian panel. Figure A2.3 compares the estimated coefficients in the three models.

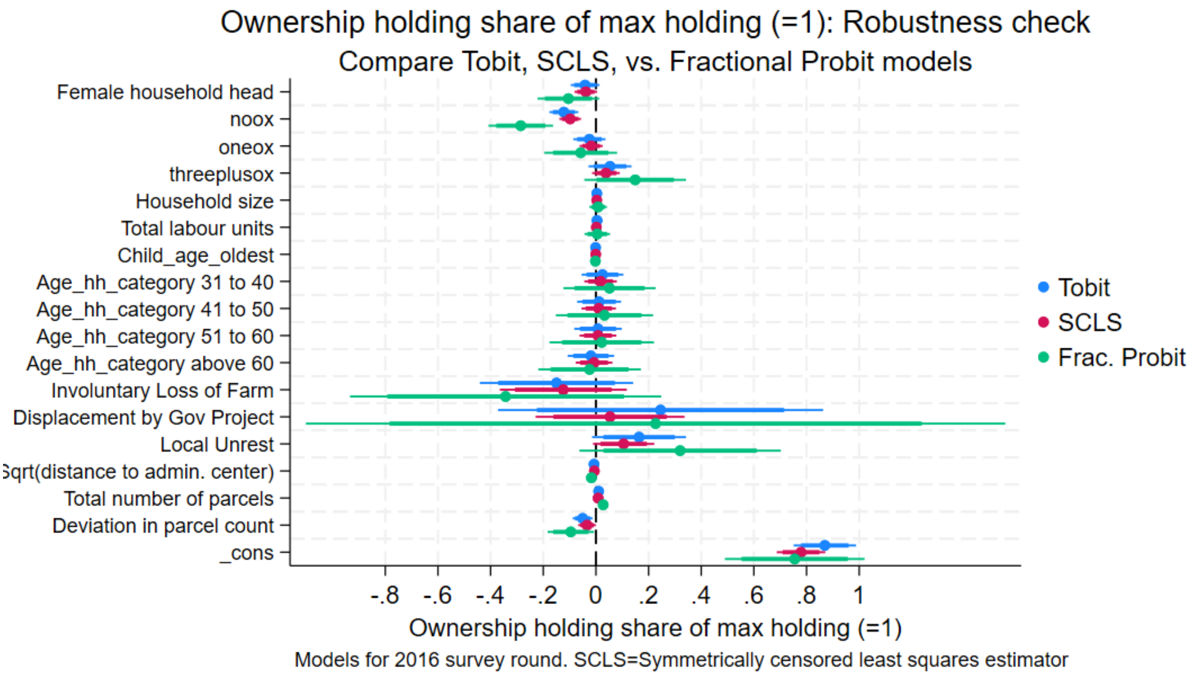

Figure A2.1. Comparing the results from the censored Tobit model with the symmetrically censored least squares (SCLS) and fractional probit estimators for 2016 data from Ethiopia.

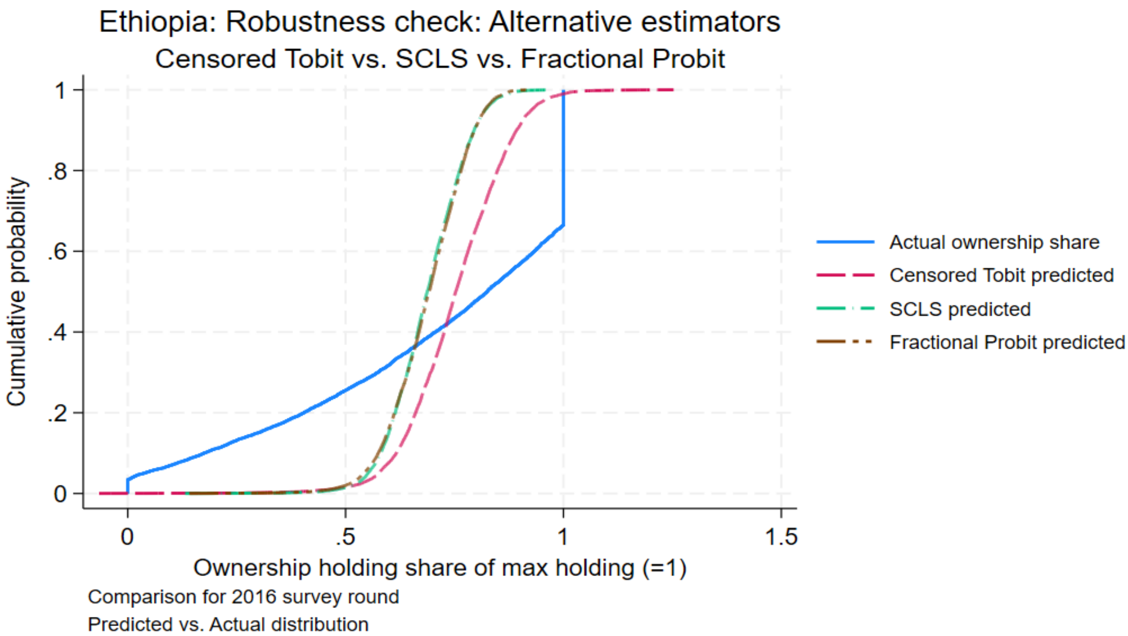

Figure A2.2. Comparing the predicted cumulative distributions versus the actual distribution of ownership holding shares of max holding from censored Tobit vs. SCLS vs. fractional probit models for 2012 data from Ethiopia.

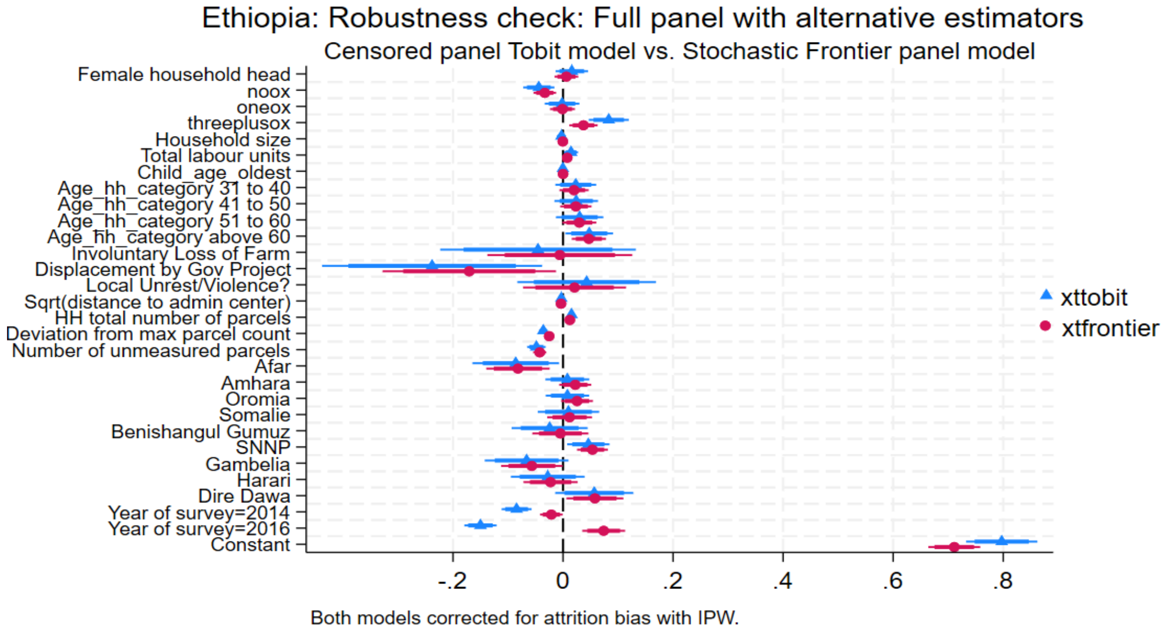

Figure A2.3. Comparing coefficients in alternative models for full Ethiopia panel (2012, 2014, 2016) from panel censored Tobit and panel stochastic frontier models.

Figure A2.3 shows that the estimated coefficients by the two models are similar for most of the variables. The exception is the dummies for survey rounds. Noteworthy are the very narrow confidence intervals for the parcel count variables that are highly significant compared to the Displacement by Government Project variable, which also was highly significant but with much wider confidence intervals. Table A1.1 shows the very low frequency of such Displacement by Government Projects in the data. It means that this variable only affects a small fraction of the sample and cannot explain much for the large share of the sample with small ownership holding shares.

Figure A2.4 presents the cumulate predicted ownership holding shares of the panel censored Tobit and the panel stochastic frontier models. The predicted distributions from the two models closely overlap, and both models show a poor fit with the actual ownership holding distributions. We see that both models produce biased estimates of the ownership holding shares, with an upward bias for those with actual shares below about 0.42 and a downward bias for almost all observations above about 0.42.

These results indicate that our approach of using the within-household maximum holding is the best among the imperfect predictors we have as a measure of household farm size as it is the least likely to suffer from omitted plot/parcel bias. We also see a very tiny fraction of the sample is predicted to have ownership holding shares that are larger than one, but these models are not likely to be good at correcting such cases, another reason for relying on the maximum within-household farm size in our follow-up analyses.

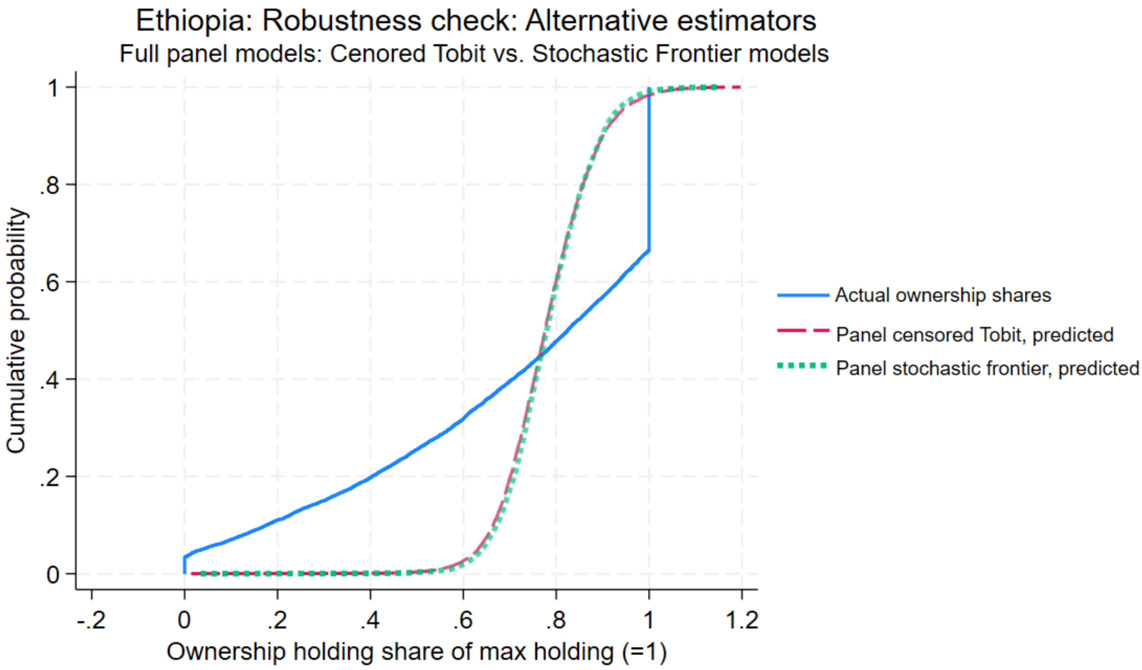

Figure A2.4. Robustness check for alternative estimators: Predicted conditional means cumulative distributions of ownership holding shares from panel censored Tobit, fractional probit, and panel stochastic frontier models versus actual ownership holding shares.

Although we cannot claim that the maximum ownership holding size across survey rounds represents an upper bound of household farm size, our predictions support the view that it represents an upper bound for the large majority of households. And it represents the best estimate we have of their ownership holding size.

Appendix 3. Exploring the within-household average variation in ownership holding shares across years and across countries

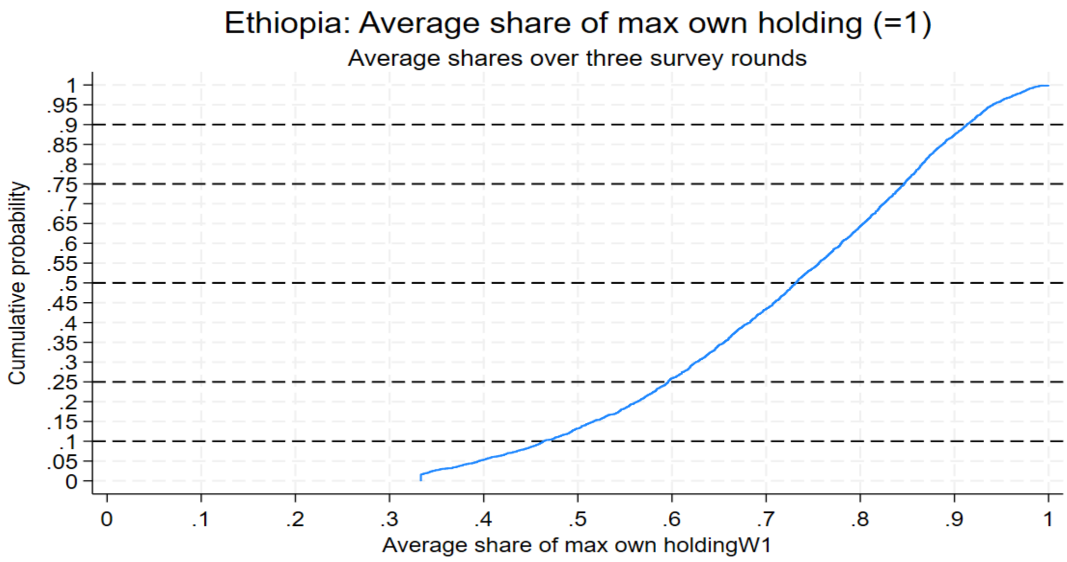

Figure A3.1. Ethiopia: Variation in within-household average ownership holding shares of max own holding over three survey rounds.

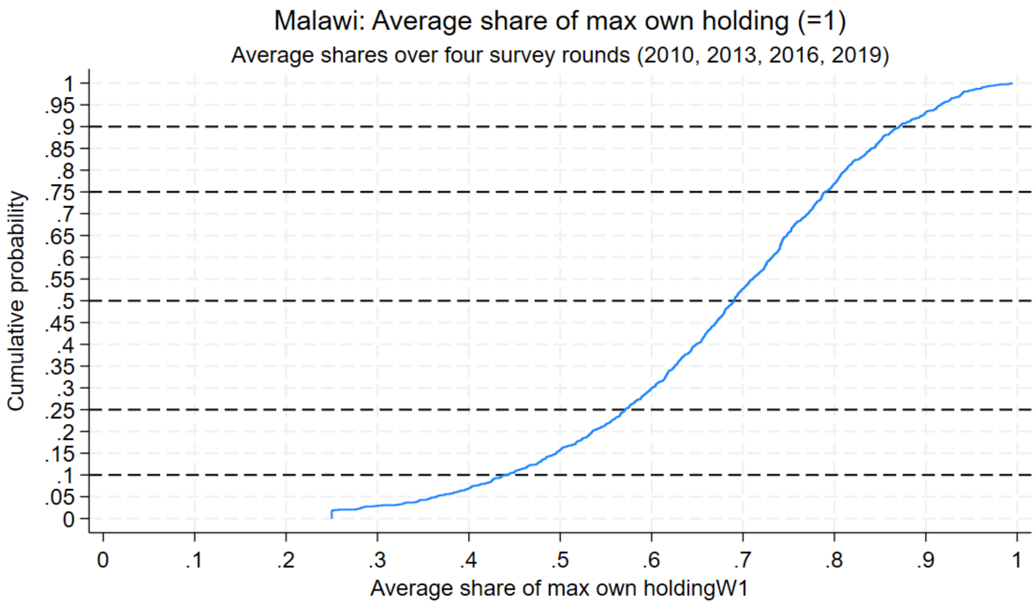

Figure A3.2. Malawi: Variation in within-household average ownership holding shares of max own holding over four survey rounds.

Figures A3.1 and A3.2 show the average ownership holding shares of maximum within-household farm sizes across survey round in each country. It shows that the median household has an average holding share that is just above 0.7 in Ethiopia and just below 0.7 in Malawi, demonstrating the similarity across the two countries. The graphs illustrate large variations in reported farm sizes within a period of 4 years (2012-2016) in Ethiopia and within 9 years in Malawi (2010-2019).

Appendix 4. Assessment of the quality of parcel ID matching across survey rounds for sampled households across the ownership holding share distributions

The enumerators collecting the LSMS-ISA parcel- and plot-level data were instructed to use the Parcel IDs from the previous survey round when collecting these data, and ideally, this should facilitate matching of parcels across years by Parcel ID even though the GPS location coordinates are not publicly available. To inspect the quality of parcel matching across survey rounds within households, we inspected the measured parcel sizes by Parcel ID at different levels of the ownership holding shares of max holding in each survey round in each country. We do this for household located as p10, p25, and p50 in each survey round and for households identified as p10, p25, p50, p75, and p90 for the average ownership holding shares presented in Figures A3.1 and A3.2. Tables A4.1 (Ethiopia) and A4.2 (Malawi) present the detailed measured parcel sizes by parcel ID for each of these households across survey rounds. The tables demonstrate generally substantial discrepancies in the GPS-measured parcel sizes across survey rounds for most households. This is another clear indication that the instructed parcel and plot matching across survey rounds was not followed in the implementation of the parcel-level surveys.

Table A4.1. Ethiopia: Parcel size (ha) distribution over the years for selected households along the ownership share distribution

hh_id	Location on distribution	ParcelID	2012	2014	2016
01020100207012	P10, 2012	1	.03531	.02151	.255428
04081702801107	P25,2012	1	.361255	.4096	3.38352
		2	.448635	.65701	
		3	1.120193		.9914
		4		.19275	
		5		.65339	
07200500604040	P50,2012	1	1.274763	.94106	.991675
		2	.25881	0	.848614
		3	.20699		.400103
		4		1.0054	
03050701204006	P10,2014	1	1.145226	.01092	.01252
		2	.658233	.11957	.0585
		3	.129931	.0588	.40258
		4	.600432	.40258	.11957
		5		.06656	
		6		0	
		7			.040475
		9			.2396
		10			.09094
13010100303128	P25,2014	1	.043068	.03031	.04965
		2	.006755	.006375	.00617
		3	.075441	.11792	.11452
		4	.115415	.08084	.12645
		5			.11948
02010400201137	P50,2014	1	.330417	.275178	.380232
		2	.157659	.145083	.063135
		3	.624776	.437581	.427595
		4	.47451	.513755	1.2965
		5	.12111		1.00025
07010800701152	P10,2016	1	1.127045	.757206	.193308
15010201201021	P25,2016	1	.01483	.0307	.023995
		2	.16729	.45674	.1332
		3	.21436	.12109	.18042

		4	.34333	.23935	.07621
04060501004048	P50,2016	1	.074218	.084337	.060336
		2	.089719	.037044	.087216
		3	.13103	.094863	.1514
		4	.226919	.18101	.14695
		5	.097549	.096308	.11741
		6	.06016	.059813	.065868
		7	.147589	.14852	.215773
		8	.176264	.1318	
		9	.131568	0	.20233
		10	.18042	.095252	.093741
		11	.08912	.076888	.071824
		12	.065786	.095536	.14823
		13	.144223	.244372	.06892
		14	.094178	0	.24116
		15	.333414	.224287	
		16	0	.161251	.181
		17	.131849		
		18			.132154
		19			.08369
07010800701003	P10,2012/16	1	.603355	.188991	.047424
04121600101091	P25,2012/16	1	.05225		.03852
		2	.94422	.46267	.17507
		3		.1084	
06020100802255	P50,2012/16	1	.08215	.1509	.14939
		2	1.045115	0	.56585
		3	.705625		.531945
		4	.79178		1.21765
		5	.26746		.03923
		6	.37602	0	.13975
		7		0	.03738
07041700202038	P75,2012/16	1	.32043	.288918	.241967
		2	.06641	.112335	.096121
		3	.09335	.081413	.057614
01010301804004	P90,2012/16	1	.654269	.670395	.579211
		2	.141831	.160225	.065898
		3		.00042	.004511
		4			.079349

Table 3.3. Malawi: Parcel size (ha) distribution over the years for selected households along the ownership share distribution

hh_id	Location on distribution	Plot ID	2010	2013	2016	2019
206157350118	P10,2010	R1	.3520842	.8093889	.3520842	.5220559
		R2		.4046945		1.392149
		R3		.4451639		.0364225
		R4		.1699717		
209066050172	P25,2010	R1	.3278025	.5908539	.3520842	.1821125
		R2	.287333	.295427	.6070417	.0687981
		R3	.1699717	.1375961		.0364225
		R4	.2468636	.218535		.2306758
303050820109	P50,2010	R1	.513962	.4856334	.9793606	.3723189
		R2	.1092675	.364225		
		R3	.0202347			
209065240180	P75,2010	R1	.3358964	.3278025	.1740186	.1740186
		R2	.368272	.3925536		
302123070520	P10,2013	R1	.1618778	.2023472	4.046945	.0364225
		R2	.1618778	.1618778		
		R3		.0809389		
204073980148	P25,2013	R1	.4451639	.4856334	.0526103	.1295022
		R2	.2832861	.2428167	.2428167	.295427
		R3	.1416431	.364225	.0526103	.0121408
		R4			.0283286	
		R5			.1133144	
		R6			.1416431	
		R7			.1537839	
		R8			.0930797	
204074810130	P50,2013	R1	.4532578	.4451639	.2023472	.0687981
		R2	.214488			.0323756
208044050380	P75,2013	R1	.586807	.586807	.9024686	
		R2	.1052206	.29138	.0283286	.8417645
		R3	.0404694			
301030740091	P10,2016	R1	1.092675	.1214083	.1092675	.3399433
		R2	.1618778	1.4569		
206144770048	P25,2016	R1	1.214083	.5665722	.6879806	.1092675
		R2	.0971267	.1618778	.0768919	.0687981

		R3	.1942533	.1011736	.0849858	
		R4	.1902064			
306022350100	P50,2016	R1	.295427	.2428167	.1821125	.1578308
		R2	.1375961			
204073980185	P10,2019	R1	.7810603	.1335492	.2347228	.1335492
		R2	.7203561	.1537839	.0607042	
		R3	.0971267	.4613517	.072845	
		R4	.0971267		.3237556	
		R5			.0364225	
		R6			.0607042	
		R7			.0526103	
		R8			.0445164	
		R9			.0809389	
		R10			.1335492	
		R11			.149737	
204040320075	P25,2019	R1	.1821125	.4937272	.8417645	.3601781
		R2	.0930797	.0647511	.0161878	.2873331
		R3	.1699717			
		R4	.4249292			
301012800179	P50,2019	R1	.3601781	.2428167	.0566572	.2832861
310043120109	P10,2010/19	R1	.4532578	.2468636	.2832861	.149737
		R2	.1699717	.6960745		
		R3		.4492109		
310017490140	P25,2010/19	R1	.3075678	1.011736	.3844597	1.31121
		R2	.2549575	.6070417	.3399433	
		R3		.8903278		
208031560034	P50,2010/19	R1	.8498583	.4856334	.8498583	.4573047
		R2	.0809389	.3035208		
101034140108	P75,2010/19	R1	.4977742	.4046945	.4856333	.1335492
		R2		.2023472		
207053720020	P90,2010/19	R1	.8093889	.4046945	.3075678	.4046945
		R2	.6636989	.8093889		.0768919
		R3	.3844597	.2428167		
		R4		.4046945		