ICAE 2024
32nd International Conference of Agricultural Economists
2-7 August 2024 I New Delhi I India

Navigating the Measurement Frontier: New Insights into Small Farm Realities

Hope Michelson[1]

1: University of Illinois, Urbana-Champaign
Corresponding author email: hopecm@illinois.edu

Abstract

Measurement is not only a way of describing complex realities; it can also transform them by influencing policies and interventions. We are privileged to live in a thrilling era of measurement innovation: new and better methods to deploy, and new ways of adapting familiar and proven apparatus to new problems and contexts. This paper explores how new measurement strategies are providing fresh insights into the circumstances of small-farm household worldwide and describes challenges that these techniques have yet to overcome. Because the small farm sector plays a crucial role in global food security, global value chains and rural livelihoods, understanding its conditions and dynamics is a persistent focus of policymakers and researchers. I discuss how satellite-based assessments of crop yields, tree cover, temperature, and rainfall, laboratory measures of soil and agricultural input quality, GPS-based plot area calculations, labor activity trackers, and high-frequency household surveys conducted via cellular phones are providing improved understanding of fundamental dimensions of small farms and agrarian households. I identify important gaps in what is currently measured, discuss challenges related to implementing and interpreting new measures, and argue that new measurement strategies can be combined effectively with continued sustained investment for traditional "analog measures" – the household and farm surveys that remain fundamental for data collection in low-income countries.

JEL Codes: C81, Q18, Q12, R14

IAAE

Navigating the Measurement Frontier: New Insights into Small Farm Realities

Hope Michelson
University of Illinois, Urbana-Champaign

July 24, 2024

Abstract

Measurement is not only a way of describing complex realities; it can also transform them by influencing policies and interventions. We are privileged to live in a thrilling era of measurement innovation: new and better methods to deploy, and new ways of adapting familiar and proven apparatus to new problems and contexts. This paper explores how new measurement strategies are providing fresh insights into the circumstances of small-farm household worldwide and describes challenges that these techniques have yet to overcome. Because the small farm sector plays a crucial role in global food security, global value chains and rural livelihoods, understanding its conditions and dynamics is a persistent focus of policymakers and researchers. I discuss how satellite-based assessments of crop yields, tree cover, temperature, and rainfall, laboratory measures of soil and agricultural input quality, GPS-based plot area calculations, labor activity trackers, and high-frequency household surveys conducted via cellular phones are providing improved understanding of fundamental dimensions of small farms and agrarian households. I identify important gaps in what is currently measured, discuss challenges related to implementing and interpreting new measures, and argue that new measurement strategies can be combined effectively with continued sustained investment for traditional "analog measures" – the household and farm surveys that remain fundamental for data collection in low-income countries.

Fundamental to economic theory and empirical analysis, sound traditional measurement strategies and plausible new ones extend the research frontier, enriching our understanding of problems under scrutiny and improving our address to these challenges. Because development economists rely on timely and accurate descriptions of economic growth, poverty, human capital, natural capital, and market functioning, the data we collect, and how we collate and understand them, are obviously essential. Such measurement strategies can improve not only through technological advances, but also through cost reductions without compromises in accuracy: granular insights acquired broadly, simply, and frequently. We study soil quality, crop production, land area, weather and its associated shocks and disruptions, prices, distances, infrastructure quality. We collaborate with crop scientists and agronomists; we pull data from government extension services and marketing boards. Our appetite for new and improving measures open paths to new economic insights but also brings methodological responsibilities, econometric puzzles, and interpretive complications.

This paper provides a brief review of recent advances in measures related to analysis of the small farm sector in low- and middle-income countries. I discuss some areas of small farm sector research where existing measures are limited or in need of further innovation. I stress the importance of returning to the question of why and what we measure, of grounding our measures and measurement investments in strong research questions. I discuss the responsibility of the careful evaluation of a new measure developed outside of economics but incorporated into our discipline. I conclude by working through some opportunities and puzzles that new measures present with examples from recent research.

The question

Agricultural development economists focus on the circumstances and dynamics of the agrarian sector in low- and middle-income countries. Increasing crop yields while reducing poverty and food insecurity is a long-standing goal of many low-income country governments, especially those in Sub-Saharan Africa. This goal has become increasingly urgent as countries also deal with increasing environmental and population pressures. Nonetheless, recent research suggests that crop yields in Sub-Saharan Africa are not increasing (Wollburg et al. 2024) while food insecurity is on the rise (FAO 2023). These concerning trends provide urgency and consequence to the study of long-standing questions in agricultural development.

Agricultural economists studying the small farm sector in low-income countries focus on regions and production environments often characterized by incomplete, imperfect, or missing markets. This persistent engagement with systemic market failures distinguishes us among the broader community of agricultural economics.

Agricultural economists studying the small farm sector in low-income countries also focus on the development challenges associated with agriculture and rural development: agricultural productivity and technology adoption; agricultural input and output market functioning and market access; land tenure and rural labor markets; the design and evaluation of rural and agricultural development policies. This focus on the agricultural sector distinguishes us among the community of development economists, who are more abstractly concerned with factors that contribute to economic growth and poverty reduction.

Production function estimation, a workhouse analysis used by agricultural economists to analyze production and allocative efficiencies, demonstrates the range of measures that can be relevant to our research. Production function estimation requires precise measures of agricultural output, and careful measures of all inputs into production: land area under cultivation; labor hours and costs, costs of inputs including fertilizers, seeds, and agri-chemicals; investments in machinery and other farm capital; and measures of additional factors including temperature, soil quality, rainfall quantity and timing, and market prices. Total factor productivity (TFP) analysis is another critical tool used by agricultural economists alongside production function estimation. TFP analysis assesses how efficiently inputs are being used in the production process and requires a similar set of measures: the quantity and quality of agricultural output; detailed accounting of inputs including land under cultivation, labor hours and associated costs, expenditures on fertilizers, seeds, and agri-chemicals, investments in farm machinery and other capital assets. Additionally, factors affecting productivity variation, such as temperature, soil quality, rainfall patterns, and market prices, need to be precisely quantified and factored into the analysis. Collectively, these measures enable economists to discern the efficiency gains or losses over time and identify the underlying drivers of agricultural productivity changes.

Other important research questions and their associated analytical approaches in our discipline demonstrate the value of an array of high-quality measures linking agricultural production data with measures of small farm household welfare and demographics. Among these are estimations of poverty dynamics and poverty traps; of agricultural household models linking household production and welfare; of impacts of agricultural production practices on natural capital and ecosystem services; of the effects of shocks on agricultural production and household welfare; and of the impact of agricultural policies on farm prices, production outcomes, and household welfare. New research questions in the field have also pushed the development of new measurement strategies, including women's empowerment indices (Alkire et al. 2013; Malapit et al. 2019) and small farmer resilience (Barrett et al. 2021; Upton et al. 2021).

In fact, the incomplete, missing, and imperfect markets that define the environment of small farmer decision-making requires agricultural economists to take more variables into account. For example, the nonseparability of household consumption and production in smallholder farm contexts means that a host of variables related to household consumption can impact household production decisions, and vice-versa.

It is easy to see why measurement challenges are numerous and significant for agricultural economists working in low-income countries on the small farm sector. Farms are often small and fragmented; land and labor markets are largely informal; farmers tend to still heavily rely on family labor for production and harvesting; and household expenses and market access as well as input and output prices can be characterized by important within-year and within-season variability. When farmers are surveyed for information about investment, revenues, and production, their recall can be uncertain, based on undocumented or poor record keeping with regard to production costs, expenses and sales.

Terminology

Two distinctions are helpful regarding the kinds of concepts that agricultural economists measure. To begin with, we distinguish between **observable variables** vs. **latent constructs**. Observable variables are those that can be directly measured using tools or observation: weight, distance, age. Latent constructs are concepts that are not so readily quantified – theoretical concepts that we cannot measure directly and must use proxies to assess: food security, socioeconomic status, intelligence. We use observable variables to measure latent constructs: income to proxy for socioeconomic status for example, or test scores for intelligence.

The second distinction: observable variables – including those being used to measure a latent construct - can be measured using **objective measures or subjective measures**. Objective measures tend to be transparent and reproducible: for instance, height and weight, temperature, water pH, land area. Subjective measures are based on self-reports gathered in interviews and surveys: personal perceptions, interpretations, opinions. Even so, not all survey-based measures are subjective; demographic and education data for example are objective measures often elicited using household surveys. An observable variable can be measured both objectively and subjectively.

A latent construct poses special measurement challenges related to the translation of an abstract concept into a concrete measure. Many latent constructs have multiple plausible and operationalized measures; some can be context-specific, varying across populations and time. Others may have several relevant dimensions. Poverty for example can be measured using a range of proxies: household asset wealth, household expenditure or income data, or composite indices that include some combination of assets, infrastructure, education, and health (Alkire et al. 2015). Assessing and interpreting a range of measures for a given concept poses its own interpretive opportunities and challenges. As abstract concepts, latent constructs can be characterized by a range of conceptualizations and interpretations – across researchers, across respondents, and between a researcher and respondents.

Operationalizing a latent construct means translating its abstract definition into an observable measure – either objective or subjective. Food insecurity for example can be measured based on an individual's objectively measured caloric intake (though this is costly and uncommon as a measurement practice). Food insecurity can also be measured using one of a host of subjective measures based on an individual or household's experience including the coping strategies that the household reports having employed (the reduced coping strategies inded, or rCSI), the household's reported access to and intake of food (the Food Experience Scale - FIES) or the household's reported consumption of a range of foods over a designated time period (the food consumption score – FCS – or the household dietary diversity score - HDDS). Several other latent constructs have special relevance for agricultural research in an era of environmental change: sustainability, ecosystem services, environmental degradation. These are important concepts to track and measure but they are also perceived as multi-dimensional and nuanced, and their measurement has long frustrated researchers and policymakers (Parris and Kates 2003).

Measurement error

Either way, a measurement is always an inference, and both objective and subjective measures can be characterized by error, differences between the true value of a variable and the measured value (Wooldridge 2008; Carletto et al. 2021). Measurement error is a problem for the internal validity of an analysis. Critical variables for estimation of agricultural production functions, for instance, are always troubled by inaccuracy: labor inputs by plot or crop; agricultural wages; labor quality; input quantity and quality; and measurement error remains an important source of bias in agricultural production functions.

Random measurement error, or "classical" measurement error as it is sometimes called, is noise – mean zero and uncorrelated with true values (in a regression model framework - uncorrelated with true dependent and independent variable values and uncorrelated with the equation error). Classical measurement error can result from factors including mis-calibrated instruments (objective measures), recording errors, or idiosyncratic reporting mistakes or variable question interpretation by respondents (subjective measures). Classical measurement error in a regressor biases the estimator toward zero (attenuation bias) and in the dependent variable increases standard errors. The presence of random measurement error increases statistical noise, pushing up the sample sizes required for analyses and increasing costs. In contrast, nonrandom, non-classical measurement error is measurement error that exhibits some relationship between, for example, the errors in measurement and the variable's true value. Biases in parameter estimation due to non-classical measurement error can mislead researchers and policymakers. Recent analyses of the implications of non-classical measurement error in agricultural data and analyses include Abay et al. (2019) and Carletto et al. (2013).

Improved measurement can ex ante reduce both random and nonrandom measurement errors (rather than using ex post statistical techniques to address the problem). Improving measurement to reduce error means increasing the precision and accuracy of our objective, instrument-based measures or improving the clarity and quality of our survey-based measures. Defining the measurement error of a measure of a latent construct involves assessing the reliability and suitability of its observable proxies.

Recent methodological advances to measure small farm realities and dynamics

Measurement has experienced profound changes over time, with gradual shifts punctuated by periods of significant evolution: improvements in time-keeping and associated ability to determine longitude transformed navigation in the 18th century; the French Revolution origins of the metric system; the international adoption of the metric system over the course of the 19th century and associated steps forward in standardizing modes of measurement (Klein 1974 provides a history of measurement and measures related to length, area, volume, weight, temperature, and position among others).

The current era has been characterized by new and more accurate measurement tools, and declining costs for measurement and data storage. Remote sensing, mobile technology, tablet computers and cell phones, and new sensors have generated a wealth of real-time, high-frequency, and spatially disaggregated data. These technologies have revolutionized the study of important variables including soil quality and soil moisture, temperature, rainfall, and population. In recent years, agricultural economists focusing on small-scale farming have incorporated many of these new measures into their analyses. Some work has combined new measures of variables not previously observable to farmers with randomized controlled trials or field experiments to study the way that farmers value and use new information about their soil quality, agricultural input quality, or land area, among others.

Remote sensing

As costs of computational power and satellite imagery access have declined, agricultural economists studying small farms have begun incorporating a range of remotely sensed data into their analyses at various spatial and temporal scales. These remotely sensed data, acquired from instruments or sensors aboard satellites, aircrafts, or drones, can provide information about crop types and land use, production volume, irrigation infrastructure and use, tree cover and biomass, plant density, temperature, rainfall, soil moisture, growing degree days and extreme heat degree days, deforestation, and land degradation (Weiss et al. 2020 and Auffhammer et al. 2013 provide reviews). Such variables are modeled from measured quantities like radiance or reflectance values, captured by sensors in various spectral bands. These data are processed using calibrated algorithms to derive measures of growing conditions and land use, facilitating research and policy decision-making in agricultural management, resource allocation, and environmental sustainability.

Because remotely sensed measures can achieve significant spatial coverage at relatively low cost, they allow researchers to monitor locations that can be difficult and expensive to reach otherwise (Wheeler and von Braun 2013). Remotely sensed measures can also refresh data with unprecedented frequency, keeping track weekly or even daily growing conditions and outcomes. Remote-sensing indices including the normalized difference vegetation index (NDVI) and measures of solar-induced florescence can be used to monitor crop stress and estimate crop production. Handheld global positioning systems (GPS) devices have become significantly more affordable and easier to use in recent years. The costs and complexity therefore of mapping plots, of adding latitude and longitude coordinates to other information about houses and plots, and of linking household with remote sensing data have also decreased considerably.

Geospatial data on environmental conditions have allowed rapidly updated spatially granular measures of conditions relevant to agricultural production, harvesting, and marketing: temperature, rainfall, and soil moisture, and have consequently reduced omitted variable bias in analysis of farmer decision-making and investment. A special benefit of fresh and extensive geospatial data on environmental conditions is improved awareness of the incidence and severity of pest and weather shocks. Information on these unexpected hazards for households, villages, regions open new methods to address longstanding questions about the near- and longer-term effects of shocks can have direct bearing on efforts to consequences for agricultural production, infrastructure and markets, and household food security, including design, implementation, and evaluation of social safety nets and humanitarian aid allocations (Lang et al. 2020; Quinn et al. 2018; Varshney et al. 2015). Improved and more temporally and spatially granular and accurate measures of agricultural risks has also been essential to the design and evaluation of index insurance for small farmers (Carter et al. 2017; Benami et al. 2021; Rosenzweig and Udry 2014).

These data have also created new research opportunities at the intersection of agriculture and the environment in low-income countries; with studies exploiting the affordability and availability of remotely sensed data on variables including wildfires and crop fire incidence and extent, crop management, irrigation infrastructure and water use, to study how farming and farmers respond to environmental regulation and policy and how farming practices impact local environmental conditions.

Satellite-based crop production estimation.

Crop-cuts are the gold-standard method for estimating crop production and yields in small farms. To implement crop-cutting yield measures, mature crops on randomly determined areas of a selected plot are harvested, dried, and weighed by researchers to estimate yields per hectare. In such studies researchers have shown that farmer reports tend to exhibit non-classical measurement errors in comparison with crop cut based measures, with reporting errors often correlated with the size of a field and its associated production (Desiere and Jolliffe 2018; Gourlay et al. 2019). Satellite -based yield estimations have shown effectiveness for regions where crop production is large scale and relatively uniform (e.g. United States, Europe, Brazil, China, and Australia) (Cai et al. 2019; Guan et al. 2016; Kern et al. 2018; Panek and Gozowski 2020; Schwalbert et al. 2020). These methods rely on NDVI to model biomass or on measurements of solar-induced fluorescence (SIF) of chlorophyll as a proxy for plant photosynthetic activity. Promising research is underway to apply these remote-sensing based crop production models to small farms in low-income countries (Burke and Lobell 2017; Lobell et al. 2019). There are significant challenges, however, for application of satellite-based crop production measures to small farms with irregular plot boundaries; intercropping (Wineman et al. 2019; Lobell et al. 2019; Tamim et al 2024); and deficient geo-referenced production and yield data (from crop-cuts or self-reports adjusted based on crop-cut data) for ground-truthing and model calibration (Burke and Lobell 2017). Note that the reliability of measures of soil quality based on remote sensing are similarly dependent on the degree of spatial variation in relevant measures; because we tend to lack good ground data on soil parameters at sufficiently granular spatial scale across low income countries, the models remain limited and their accuracy unestablished (Gourlay et al., 2017; Kosmowski, et al. 2020); finally, a mismatch persists between the small scale of most farming operations in many low income countries and the spatial resolution of affordable geospatial data on soil properties.

GPS-based plot area measures

Land area can be objectively measured using compass-and-rope (Carletto et al. 2017), an accurate but costly and time-consuming process that remains the gold-standard method. For many years economists working with small farmers have instead relied on reports from farmers of the sizes of their plots. In regions without land registration and without taxation based on land area and therefore without independent surveyors measuring plot boundaries and associated land area, however, farmer reports may prove subjective. Plot area measurement based on enumerator plot mapping with handheld GPS units has been shown to compare favorably with compass-and-rope measures (relative to farmer reports) and to correct errors and exaggerations in farmer reports (Cohen 2019; Dillon et al. 2019; Carletto et al. 2016). As a result, GPS measurements have recently been integrated into several agricultural household surveys for the measurement of household parcels and plots. This method does require enumerator training and requires a site visit to all plots to be measured, which can add cost to the survey depending on the distance of the plots from the household and from each other. Comparisons of farmer-reported land area versus GPS-based measures suggest significant differences (Carletto, Gourlay, and Winters 2015; Diillon et al. 2019; Abay et al. 2019; Abay et al. 2023), attributable to farmer reporting errors and enumerator errors, respectively. As I discuss in the final section of the paper, an accurate objective measure may not always be the preferred choice for a researcher depending on his or her question. Research questions concerned with farmer decision-making and behavior, for example, may glean as much insight from farmers' subjective beliefs about their land area (and yields) as from an objective area measure. In combination, the two measures together can be very informative about farmer beliefs, investments, and production.

Remote sensing and poverty measures

Researchers have begun exploring the value of remotely sensed imagery in combination with existing survey data and machine learning techniques to train models that can predict poverty at a range of spatial scales (Yeh et al. 2020; Ayush et al. 2020; Chi et al. 2022). Jean et al. (2016) for example use demographic health survey (DHS) survey data, Living Standards Measurement Study survey data from the World Bank, and daytime satellite imagery to infer poverty rates in areas where ground-level data is sparse or unreliable. This new approach is reminiscent of small area poverty estimation work of the early 2000s which combined census data and detailed household surveys with techniques from small area statistics to address the same problem: a lack of high quality spatially disaggregated data on poverty in low-income countries (Elbers et al. 2003; Alderman et al. 2002). In fact, Van der Weide et al. (2021), comparing small area poverty estimates in Malawi based on census data, household expenditure survey data, and estimates based on remote sensing data, find that these two approaches provide similar insights about the spatial patterns of poverty.

Poverty estimation based on daytime satellite imagery - like the analysis of nightlights from nighttime satellite imagery – is a method to extract socioeconomic information from remotely-sensed data. Poverty estimation from daytime imagery produces more spatially granular estimates of poverty than analyses of nightlights from satellite images, which has been used to compare economic activity across countries and time but lacks sufficient precision in the signal to quantify within-country variation in most low-income countries (Henderson et al. 2012). Research that uses mobile phone call data records to estimate poverty within countries has been hampered by legal and logistical complications for researchers accessing proprietary data held by telecommunications companies (Steele et al. 2017).

Moreover, use of satellite imagery to estimate spatially granular rates of poverty has limitations. Estimates based on satellite imagery are likely to provide an inaccurately static measure of regional poverty if significant variables picked up by the model in the training data are relatively fixed in time (infrastructure quality and access, housing quality including roof material, terrain and irrigation for example). A satellite measure may fail to capture important characteristics of poverty including within-year seasonal changes, spatial income disparities or social vulnerabilities. Furthermore, data privacy concerns and as yet limited geo-referenced high-quality asset, expenditure, and consumption data to train prediction models limit broader application (Ayush et al. 2021) and remote sensing data used in concert with survey data can also be characterized by measurement error from sources: miscalibration in the measurement technology, misspecification of the modeling converting the signal from the sensor or satellite to the measured variable, and the inferred spatial or temporal resolution of the data. Despite such limitations, satellite-based poverty estimation shows promise in complementing traditional survey methods.

Lab-based measures

Soil quality

Though soil quality and input quality are usually not directly observable by researchers or farmers, they are crucial determinants of agricultural productivity (Barrett et al. 2009) and may influence farmer learning about the suitability and profitability of new technologies and techniques (Tjernstrom et al. 2017). Economists have long speculated that the significant heterogeneity in production outcomes within and across villages in SSA for example could be at least in part attributable to heterogeneity in soil quality (Gollin and Udry 2021), and agricultural economists studying the small farm sector have begun to use laboratory measures of soil nutrients to characterize spatial patterns of soil quality, to study correlations of soil quality with farmer management and environmental factors (Berazneva et al. 2019; Marenya and Barrett 2009; Gourlay and Kilic 2023), to analyze the relationship between farmer beliefs about soil quality and the objectively measured soil pH, macro and micro nutrients (Berazneva et al. 2018), and to characterize relationships among soil quality and other household socioeconomic variables (Barrett and Bevis 2015). We note that much of the research comparing farmer soil reporting of soil characteristics and lab-based measures find that the lab measured quality variables are not able to explain much of the variation in the farmer's own assessments (Berazneva et al. 2018; Kosmowski, et al. 2020). Some recent research has studied the value of lab-based soil quality measures and associated management recommendations to farmers (Harou et al. 2022; Brezaneva et al. 2023; Tamim et al. 2024). Soil quality is multidimensional, and the literature has not yet converged on the right summary indicator for quality yet, frustrating data integration and comparison across studies. A final note – spectrometers – which offer the possibility of future in situ measures of a handful of important soil characteristics including organic carbon and pH and which seem to compare well with measures from lab-based soil analyses (Kosmowski et al. 2020), the units remain costly though their prices are beginning to decline, offering exciting prospects for soil measures in the coming years.

One paradox deserves special attention as it demonstrates the way that new measures can shed light on persistent research questions in the small farm sector: the long-observed inverse relationship between crop yields and cropped area has significant implications for policy related to support for small farms in low-income countries (Barrett 1996; Rios and Shivley 2005; Sen 1962; Deininger et al. 2022). One longstanding explanation for the relationship has long been that it is attributable to omitted variables: specifically, that soil quality – strongly correlated with productivity – is better on smaller farms relative to larger farms (Benjamin 1995) because of differences in input and cultivation practices. However, analyses incorporating plot-level soil quality measures indicate that soil quality may not in fact be the answer. For example, using good soil quality data at the plot level with multiple plots per household, Barrett et al. (2009) find that soil quality measurements do not appear to account for the relationship, suggesting that a better explanation may involve factor market imperfections (Foster and Rosenzweig 2022) or measurement error (Lamb 2003; Bevis and Barrett 2020; Abay et al. 2019).

Non-labor agricultural input quality

Researchers have also begun to use lab-based measures to study the quality of agricultural inputs available for purchase by small farmers: the content of nutrients in mineral fertilizer (Hoel et al. 2024; Michelson et al. 2023; Michelson et al. 2021; Sanabria et al. 2018) and the actual amount of active ingredient in market-available glyphosate herbicide (Ashour et al. 2019). Low quality inputs not only have consequences for agricultural productivity but impact farmer decisions about using them in sufficient quantities or even at all (Assima et al. 2017; Bold et al. 2017). A focus on non-labor agricultural input quality also invites a research focus into the functioning and operations of input supply chains, an important area where agricultural economists working in low-income countries are beginning to extend their work (Naugler et al. 2024; Dar et al. 2024;

Fertilizer (most tests documented in the literature to date are assessments of the nutrient content of urea, a single-nutrient fertilizer 46% nitrogen by weight, widely used by small farmers) has been shown in this work (summarized in Michelson et al. 2023) to be reliably of good quality in these studies, with nutrient content consistent with the manufacturing standard. Glyphosate fertilizer has been found to exhibit widespread quality problems; prone to dilution with water Ashour et al. (2019) find that one third of the glyphosate samples they acquired across markets in Uganda were missing at least 25% of the active ingredient; Haggblade et al. (2021) also find many unregistered and counterfeit glyphosate pesticides in West Africa, bottles missing an average of 8-10% of the active ingredient. Seed quality has many dimensions including germination rate, disease presence, and seedling vigor; observable features of seed performance that farmers also assess in the field. Researchers are beginning to document and quantify these dimensions of seed quality (Barriga and Fiala 2020; Bold et al. 2017), though more attention has been on assessing varietal purity –

whether a seed is in fact the variety that it is advertised to be (more on this below) - especially as the costs of DNA sequencing have come down in recent years.

Researchers have collected both lab-based measures and farmer perceptions of quality for these inputs and have documented that farmer report beliefs about input quality inconsistent with lab-based analysis. For example, farmers widely report that the quality of local-market urea is poor when laboratory results indicate that it consistently meets specified standards (see Sanabria et al. 2017 and Michelson et al. 2023 for a discussion of this point and presentation of the evidence). Perceptions about glyphosate, however, present a different case. Evidence has found widespread quality problems with this herbicide, but also that farmer may perceive its quality to be worse than it actually is. Across markets in Uganda, Ashour et al. (2019) find evidence of a correlation between the documented quality problems in the market and farmer beliefs about quality in that area, but show that these beliefs are often locally inaccurate, only partially accounting for real differences in quality.

DNA fingerprinting

Agricultural economics research has begun to use DNA fingerprinting to identify specific varieties of crops under cultivation (Stevenson et al. 2018; Yigezu et al. 2018), using seed or a leaf-cuttings from growing plants for lab analysis. These DNA profiling techniques can be used for analytical and varietal purity testing of seed lots. Work of this sort has been conducted by CGIAR as part of their efforts to document the scope and impact of their decades of research on small farm crop adoption and productivity (Stevenson et al. 2018). Research has focused so far on varietal identification in cassava (Maredia et al. 2016; Wossen et al. 2019), sweet potato (Kosmowski et al. 2018), and bean (Rabbi et al. 2015). Fingerprinting is straightforward for clonally propagated crops like sweet potato and cassava and in self-pollinating crops like beans, but considerably more complex for open pollinated crops like maize (Bohr et al. 2024). Varietal identification based on DNA analysis in maize is more complex as pollination and hybrid crossing occurs in the field. Maize DNA fingerprinting requires a genetic reference library to identify and match tested samples with true breeder varieties; building and validating this kind archive can be expensive and time consuming, and it remains a work in progress for many crops (Warburton et al. 2002).

Research findings on varietal identification that includes both lab analysis of farm and field samples and farmer self-reports has revealed significant discrepancies crop varieties between what farmers believe they are growing and the actual strains under cultivation For example, Wossen et al. (2017) find that 28% of the 2,500 nationally-representative cassava farmers in Nigeria assumed they were growing local varieties when they were actually growing genetically improved plant material, while 13 percent thought the reverse – that they were growing improved varieties when in fact they were not. Wossen et al. show that this misreporting is non-classical error, correlated with a handful of farmer observables including education level and higher self-reported access to more sources of information about crop varieties. Bohr et al. (2024) document similar patterns among maize growers in Ethiopia. These apparently widespread misperceptions underscore an important theme discussed further the final section: farmers' beliefs can offer researchers valuable insights, even when they are wrong.

Output market crop quality measures

Marketed crop quality is relevant for understanding prices, market structure, and market segmentation. Some research into the small farm sector has begun to use new methods including laboratory measures of moisture and aflatoxin to measure quality of marketed agricultural production including maize in Uganda (Bold et al. 2022) and wheat in Ethiopia (Do Nascimento Miguel 2024), watermelons in China (Bai et al. 2021), milk in Vietnam (Saenger et al. 2014) and groundnuts in Ghana (Magnan et al. 2021). This research is related to the implementation and evaluation of policies designed to incentivize farmers to invest in higher quality production or post-harvest practices and also to research into contractual design and compliance in agricultural value chains. A challenge that this work faces is that quality in output markets is multi-dimensional and, for many attributes of a crop (color, damage, size), somewhat subjective. Not all relevant dimensions of quality have objective measures. Analysis of discrepancies among transacting parties and dynamic change in the definition and enforcement of quality attributes are interesting areas for further research.

Labor activity trackers

Accurate measurement of labor hours and labor productivity has lagged other advances, despite the fact that labor is a critical component of understanding productivity, allocation, investment, adoption. Household surveys remain the primary method for the collection of data about labor in agricultural production. Some recent research has tested higher frequency labor modules via cell phones (Arthi et al. 2018). When Akoguan et al. (2020) used wearable accelerometers (a Fitbit is an example) to track activity and physical effort of sugarcane cutters in Nigeria, they found that measures based on the accelerometers correlate well with firm data on worker production, and they assert that careful measures of labor effort can provide insights into the relationship between effort and productivity, or in cases where productivity is not observed, provide a valid proxy for productivity. This finding has special relevance in analysis of farm work without a market wage, including tasks like collecting water or other activities associated with subsistence agriculture.

Advances in survey methodology

New data sources based on ICT, sensors, laboratory methods, and satellites have special advantages when combined with household farm surveys, which continue to constitute the cornerstone of agricultural data collection in low-income countries. These surveys remain critical not only for validation and for ground-truthing other methods, but also for targeting, for generating essential agricultural statistics, for measuring spatial variation and patterns over time, and for providing the substrate for research into the fundamental questions of agricultural economics.

Healthy and useful empirical research must consistently scrutinize its own methods and their development. Such assessment and innovation have been underway in household survey design focused on the small farm sector. Researchers have focused on innovations including: the relevant scale and method for measuring a range of critical household and agricultural variables including plot level analysis; gender-disaggregated responses to labor and asset ownership; and investment in panel surveys linking household and agricultural data. For example, the frontier of work in recent years using multiple methods to measure the same quantity has been the World Bank Living Standards Measurement Survey team, whose Integrated Surveys on Agriculture initiative (which began in 2008) integrated detailed, high quality, plot-level agricultural data with household-level data, providing fresh insights into rural households and livelihoods and rural sector conditions in eight countries (Carletto et al., 2010). As a part of this work, the LSMS-ISA team has taken the lead to conduct careful comparisons of objective (generally a "gold-standard" method) and subjective (farmer reported) measures – much of this work discussed above - of variables including crop production quantities (Lobell et al. 2020) and farm plot size (Carletto et al., 2013; Kilic et al., 2017; Abay et al. 2023).

A second area of research on survey methodology in low-income countries has focused on data quality, using survey experiments (many implemented by World Bank research efforts) to interrogate the quality of data produced by enumerators in questionnaires and interviews. De Weerdt et al. (2020) explore the considerable effect that survey design can have on data quality and associated analyses and discuss best practices. Work studying survey methods design has provided insights into features of a survey on data quality and reliability: question ordering, survey length and respondent fatigue, and survey respondent choice. This work has focused on labor statistics in household surveys (Bardasi et al. 2011; Kilic et al. 2020), use of production diaries to improve household crop and investment statistics (Deininger et al. 2012), the effect of recall periods on household consumption and associated welfare estimations (Beegle et al. 2012) and on farmer-reported crop production (Wollburg et al. 2021). Results of these analyses (Kilic and Moylan 2016) tend to emphasize the importance of respondent selection and the design of modules on asset and land ownership for example or labor hour modules based on recall (Beegle et al. 2012). Other papers summarizing recent literature in this space and laying out best practices include Dillon et al (2021) and Carletto et al. (2021).

Cellular-Phone based surveys

Use of cellular phone-based surveys to collect household and agricultural data is developing quickly as cellular phone networks and ownership increases in low-income countries, especially among the rural poor. Cell phone-based surveying has some limitations in agricultural data collection, however, given the relative remoteness and low (though improving) connectivity of many farm households as well as the sometimes-acute poverty and lower literacy rates among rural populations. Driven by the need to reach households safely amidst mobility restrictions, and by heightened national and international interest in monitoring household welfare during health and economic crises, the adoption of phone-based surveys accelerated during the COVID-19 pandemic. (Josephson et al. 2021; Kanyanda et al. 2021; Rudin-Rush et al. 2022). An associated methodological literature on cell phone survey best practices is also providing new insight (Gourlay et al. 2021; Brubaker et al. 2021; Kastelic et al. 2020).

Because the cost of adding rounds of cellular phone-based surveys is lower than the costs of additional rounds of in person surveys, development of cellular based surveys has been accompanied by increased collection and analysis of **High Frequency Surveys.** These surveys offer several attractive features including collecting timely, even in some cases near "real-time" data. Some of these data can help alleviate accuracy problems in annual household surveys. For example, cell phone surveys have also been used to collect high-frequency labor data (Arthi et al., 2018; Dillon, 2012) and to compare high-frequency estimates with estimates of labor hours based on recall over longer periods. These high-frequency household surveys conducted via cellular phones are changing policy impact analyses and policy targeting, but also providing new understanding into the fundamentals of small farms and small farm households.

Challenges

New measurement technologies and strategies bring new challenges. Some of these are familiar, like harmonization of scale and units, or evolving, inconsistent norms for checking the accuracy of measures incorporated from outside sources. Others are specific to these new options. I present four examples below, with reference to recent work.

Evaluating and documenting measures incorporated from other disciplines

Incorporating measures that originate in other disciplines requires special attention to and documentation of data quality as processes for validation and evaluation are not native. Especially when measurement strategies are new to agricultural economists, norms for evaluation can be unclear and evolving. Authors of studies using such new measurement strategies face a heightened responsibility to document the quality and reliability of their findings. Economists using measures from other disciplines must understand and follow measurement protocols from those disciplines (Michler and Josephson 2023; Michelson et al. 2023) including how to recognize and account for error in the measure itself. For example, laboratory measures of soils or nutrient content in agricultural inputs normally involve double testing a randomly selected subset of samples to characterize measurement error. Economists adopting the strategy would do well to take an additional step of testing samples across multiple labs, to determine measurements of quality conditional on the given laboratory. This double-checking can be crucial if the study must involve labs where quality control is not assured. When Ashour et al. (2019) analyzed results from urea fertilizer tests performed in multiple labs as a validation of their data and urea nutrient content measures for a study in Uganda, they found significant differences in the reports of nutrient quality of tested urea depending on which laboratory performed the analysis. Michelson et al. (2023) describe a similar experience for urea samples double tested in Kenya and the United States. Interrogating data quality in the analysis of soil samples should include thorough documentation of the sampling design, discussion of the laboratory protocols for sample preparation, and analysis following established USDA or FAO methods. Michler et al. (2023) work through a similar set of best practices for the use of remote sensing measures in combination with household survey data. Poets et al. (2020) present emerging protocols for conducting DNA fingerprinting for varietal identification.

When is discordance informative?

We have already discussed examples of observable variables with multiple co-existing measures: observable variables that can be measured both subjectively and objectively, including soil quality, crop variety, and plot area and latent constructs with multiple proposed plausible objective and subjective measures, such as poverty, sustainability, environmental degradation, and food insecurity. Recent innovations in measurement and collection of agricultural data, with multiple data sets deploying multiple methods to measure the same variable (LSMS-ISA), have created new opportunities for comparison and analysis across measures. In many cases, however, multiple measures of the same underlying concept are not consistent. How can we interpret such discordance? When can it be informative? And when does it merely reveal measurement error rather than useful information?

Significant variations across measures can be confusing for researchers and policymakers, especially when 'gold standard' measures have yet to be agreed upon or do not exist at all. For example, food insecurity is a latent construct with considerable relevance for agricultural economics research and associated policy, and a primary household and individual welfare outcome for agricultural economists working in low-income countries. Household food insecurity measures are used for humanitarian and development programs and are critical inputs into the status quo global system of food insecurity crisis assessment and prediction and associated humanitarian aid allocation, the Integrated Phase Classification System (IPC). A range of measures exist for food insecurity including the household hunger score (HHS), the dietary diversity score (DDS), the reduced strategies coping index (rCSI), the food insecurity experience scale (FIES), the Household Food Insecurity Access Scale (HFIAS), the Self Assessed Measure of Food Security (SAFS), and the food consumption score (FCS), with new measures still being introduced and proposed (2024). Food security is conceptualized as having four hierarchical "pillars" – availability, access, utilization and stability. This array range of extant measures is recognized as capturing related dimensions of food insecurity. Even so, this range of measure relies on either recall-based objective measures of consumption characterized by classical and non-classical measurement error, or on subjective, experience-based measures. Most extant measures of food insecurity serve as proxies for food access. Because anthropometric measures (height, weight) are shaped by other factors including the disease, environment, and infrastructure access, they can only serve as lagged proxies for patterns of food consumption.

Nonetheless, research has established that, these measures, when correlated, can provide inconsistent pictures in the cross-section of food security status of households and regions (Maxwell et al. 2014). Work is underway to explore if that discordance can be informative. Maxwell et al. (2014) for example, argue that because the current slate of measures tends to be used interchangeably by researchers and policymakers, such research and policy formulations are likely impiared by by what they term "errors of exclusion". Kim et al. (2024) work from a hypothesis that there may be useful information in the intersection of indicators: in other words, that households classified as food insecure by multiple indicators may be experiencing deprivation more severe than what is indicated by a single measurement. The top panel of Figure 1 (from Kim et al. 2024) uses data from Malawi's three rounds of LSMS to calculate the percent of the population classified as food insecure based on the rCSI, the FCS, or both. The analysis also includes an asset-based poverty classification for each household (defined as the bottom 20% of the asset distribution) and the intersection of these. Only one percent of the population is food insecure based on both the FCS and the rCSI indicators and only half of those households are also asset poor (sample pooling across rounds). Most of the households identified as food insecure based on the rCSI were not in the bottom quintile of the asset distribution, and only about half identified as food insecure based on the FCS were. The bottom panel of Figure 1 presents the households classified as food insecure in both measures, and also asset poor. Kim et al. (2024) suggest that these intersections of food insecurity measures with measures of household wealth may provide contextually-relevant insight, helping for example to distinguish between chronic and acute circumstances of food insecurity. Their study reveals an important area of future work, with relevance for methods, research, and policy.

Djima and Kilic (2024) explore a related question: when and how researchers can use self-reported variables measured with error to infer or more closely approximate "true" values. Focusing on crop production quantities, they use sorghum fields for which they have both "gold-standard" crop cut measures of production and farmer self-reported measures, along with publicly available remotely sensed plot-level data on other variables (including rainfall) to calibrate a model using machine learning to predict and adjust farmer

	% from the whole sample (28688)			
	2010	2013	2016	All rounds
Asset poor (bottom 20% quantile)	25.82%	24.42%	20.43%	23.29%
rCSI poor (if rCSI>17)	4.88%	5.67%	20.01%	11.54%
FCS poor (if FCS <35)	3.79%	1.56%	3.93%	3.54%
FCS & rCSI poor	0.64%	0.40%	1.94%	**1.17%**
rCSI & asset poor	2.29%	2.61%	6.97%	**4.36%**
FCS & asset poor	2.15%	0.95%	1.93%	1.89%
FCS & rCSI & Asset poor (Chronic FI)	0.40%	0.25%	0.97%	**0.62%**
FCS & rCSI & Asset high (Acute FI)	0.00%	0.03%	0.04%	0.02%

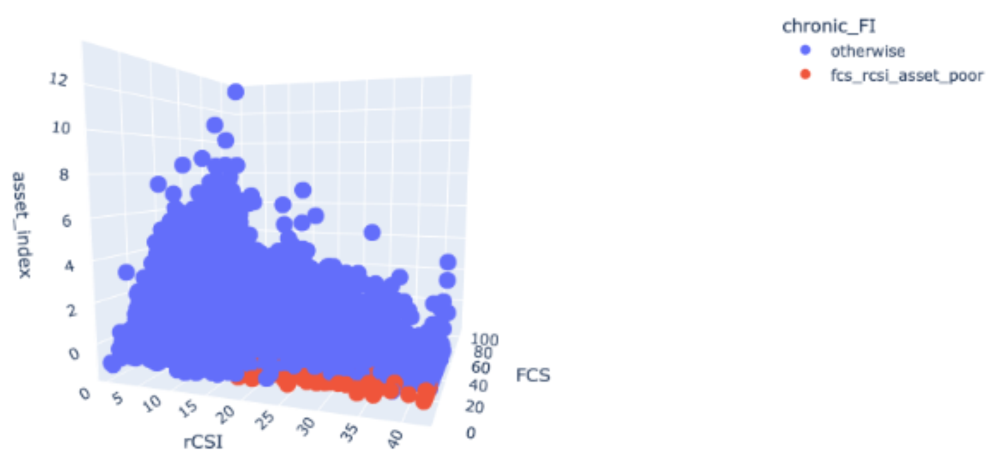

Figure 1: Table and figure from Kim et al. (2024) presenting the percent of households classified as food insecure based on the FCS and rCSI or both and the percent of households in the bottom quintile of the asset distribution who also fall into these classifications. Data is from three rounds of the Malawi LSMS IHS, pooled sample n=28688.

self-reports out of sample. They propose this method as a means of reducing the presence of non-classical measurement error that has been found to plague farmer self-reports. The authors argue that the measurement technique can be successfully implemented if one third of plots are designated for crop-cutting in the given sample (assuming relevant dimensions of variation are sufficiently similar to their Malian sample). Their analysis suggests the power of self-reported data in combination with quality objective measures of yields and geo-referencing, and it demonstrates how good georeferenced surveys can provide training data for remote sensing-based measures.

Farmer perceptions – a special case of measurement discordance?

A focus of measurement improvement in agricultural economics for the last 10-15 years has been improving the accuracy of objective measures: field size, production quantity, input quantity, input quality. As precision and accuracy of these objective measures have improved, discrepancies between what farmers report and researchers quantify have become more apparent and pronounced.

Research has begun taking these discrepancies more seriously across a range of emerging work, documenting and analyzing differences between objective findings and farmer beliefs about input quality, field size, production quantities, and soils. Abay et al. (2023) discusses how farmer misperceptions, as a form of measurement error, can provide a basis for distinguishing between farmer misperceptions and farmer misreporting of accurate beliefs. Their work demonstrates that farmer beliefs and farmer uncertainty can have important consequences for farmer investment and adoption. For example, because farmers may conclude that an improved input is a poor investment when its quality or varietal type is not what they believe it to be their perceptions can impact not only input adoption but also their openness to learning. Bulte et al. (2023) show that such uncertainty about maize seed quality impacts farmer investment, especially in complementary inputs like labor. Similarly, farmers holding incorrect beliefs about the type of seed they are growing (for example, assuming that a traditional seed is a hybrid) is likely to allocate inputs suboptimally, lowering returns to adoption as well as researchers" estimates of those returns (Wossen et al. 2019; Euler et al. 2022; Bohr et al. 2024).

Researchers are also working (Hoel et al. 2024) to test whether erroneous perceptions about input quality can be corrected, and to assess the impact on small farmer decision-making when these corrections are achieved (Maertens et al. 2024). This study relates to work by Abay et al. (2023) which provides farmers with the GPS-measured areas of their plots, and documents how these farmers amend beliefs about their plot size. This research can be a basis for better studies of the origins of and spatial patterns of misperceptions and their response to perturbation, the resistance to change over time, and consequent implications for farmer investment in technology and learning.

What is the relevant scale – digital extension

The appropriate spatial scale for measurements outlined above can vary. The relevant scale for one kind of research may not always be the relevant scale for policy, nor may it be the scale beneficial for farmers. In soil sampling, for instance, a challenge lies in deciding how frequently, and at what granularity, to collect samples to represent variability of soil properties within a field or region. This complexity also extends to price measurement, where determining the appropriate market level—whether farmgate level, local markets, regional hubs, or national trading centers —can significantly influence policy decisions and economic analyses. Addressing such challenges requires nuanced methodologies, advanced technology, and interdisciplinary collaboration to ensure that measurements accurately reflect the diverse spatial contexts inherent in the small farm sector.

Special consideration and attention should be given to the question of the relevant scale of information for farmers by researchers and by policymakers. A body of work is developing that provides information about growing conditions including weather forecasts, and information about local soil deficiencies and associated remediation recommendations and evaluates whether and how farmers use that information and the degree to which it impacts investments, profits, and yields (Harou et al. 2022; Corral et al. 2020; Fabregas et al. 2024; Beg et al. 2024; Gars et al. 2015; Burlig et al. 2024). The increasing use of cellular phones by small farmers means that this information can be provided at large scale and minimal cost. Plot-based soil testing of farmer plots found that information alone did not change farmer behavior but in combination with liquidity, farmers responded to soil limitations that were shared by many other farmers in their village rather than their plot's more idiosyncratic deficiencies (Harou et al. 2022) and work estimating farmer willingness to pay for tests of plots that share observable characteristics with their own plot suggests that farmers believe that they can learn from and benefit from tests of others' fields (Berazneva et al. 2024). Plot-level testing may not be the actionable scale for farmers given the paucity of information currently available at village or even regional scale. Nonetheless, information and associated recommendations at relatively more spatially or temporally (in the case of weather forecasting) coarse scales comes with a degree of uncertainty in the measure. Research into how to communicate that uncertainty effectively and responsibly to farmers is an important area for future work.

Conclusion

Though strong gains have been made in terms of technological innovation, in data integration, and in collection modalities, significant gaps remain in measurement of the small farm sector. For example, more methodological work is needed on livestock production and marketing, on labor, on the timing of agricultural investments, on the collection of input and output price data, and on the spatially granular characterization of agricultural risk. We need innovation to measure informal sectors of the agrarian economy: to characterize operations of traders, wholesalers, and "briefcase buyers" who operate at the farmgate. Household and farm surveys are not well designed to capture the features and dynamics of rural markets and supply chains that characterize the evolving food system in most countries. In addition, stronger efforts are required to link data through georeferencing plots and households, and innovation is urgent now in areas related to data privacy and data quality.

Because measurement can be a costly undertaking and investment in improving a measure's accuracy or reliability can be significant, research outlays for these purposes should be in the service of a clear objective. Though innovations can remedy existing data biases and reduce measurement errors, they can also introduce fresh complications and biases. Researchers must negotiate tradeoffs between costs and benefits of reduction in biases and errors. The degree to which a measure permits a new perspective, adds new information and clarity to an analysis, improves the precision of an estimation should be evaluated based on the context and the question but researchers must be vigilant and rigorous in paring measures and investment back to the research question at hand. Calculating the value of increasing accuracy can be difficult and requires clarity from the researcher about the relative importance of accuracy for targeting or for parameter estimation for example; the relative cost of errors of inclusion versus errors of exclusion. Carletto et al. (2017) offers guidance on these and other points.

Machine learning methods are a powerful tool with enormous potential to analyze the small farm sector. Even with their remarkable efficiency and computational power, these approaches are limited by the quality and representativeness of the training data. It is imperative that the training database accurately reflects the spatial and temporal variations of the target phenomena and observational conditions, accounting for inherent noise and biases to maintain the generalizability of these methods. Regardless of the methods employed to characterize agricultural systems, continual validation remains paramount. This will require continued investment in the workhorse data collection modalities of the small farm sector: agricultural censuses, household and farm surveys and nationally representative panels. New measurement strategies will provide the most insight if we combine them and future innovations with continued sustained investment for traditional "analog measures" – the household and farm surveys that remain fundamental for data collection, policy-making, and research in low-income countries.

Works Cited

Abay, K. A., Abate, G. T., Barrett, C. B., & 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., Barrett, C. B., Kilic, T., Moylan, H., Ilukor, J., & Vundru, W. D. (2023). Nonclassical measurement error and farmers' response to information treatment. *Journal of Development Economics*, *164*, 103136.

Alderman, H., Babita, M., Demombynes, G., Makhatha, N., & Özler, B. (2002). How low can you go? Combining census and survey data for mapping poverty in South Africa. *Journal of African Economies*, *11*(2), 169-200.

Alkire, S., Roche, J. M., Ballon, P., Foster, J., Santos, M. E., & Seth, S. (2015). *Multidimensional poverty measurement and analysis*. Oxford University Press, USA.

Alkire, S., Meinzen-Dick, R., Peterman, A., Quisumbing, A., Seymour, G., & Vaz, A. (2013). The women's empowerment in agriculture index. *World development*, *52*, 71-91.

Ashour, M., Gilligan, D. O., Hoel, J. B., & Karachiwalla, N. I. (2019). Do beliefs about herbicide quality correspond with actual quality in local markets? Evidence from Uganda. *The Journal of Development Studies*, *55*(6), 1285-1306.

Assima A, Haggblade S, Smale M. (2017). Counterfeit herbicides and farm productivity in Mali: a multivalued treatment approach. Feed the Future Innovation Lab Research Paper No.50. East Lansing, Michigan: Michigan State University.

Auffhammer, M., Hsiang, S., & Schlenker, W. (2013). Using weather data and climate model output in economic analyses of climate change. Review of Environmental Economics and Policy, 6.

Ayush, K., Uzkent, B., Tanmay, K., Burke, M., Lobell, D., & Ermon, S. (2021, May). Efficient poverty mapping from high resolution remote sensing images. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 35, No. 1, pp. 12-20).

Bai, J. et al. (2021). Melons as lemons: Asymmetric information, consumer learning and seller reputation.

Bardasi, E., Beegle, K., Dillon, A., & Serneels, P. (2011). Do labor statistics depend on how and to whom the questions are asked? Results from a survey experiment in Tanzania. *The World Bank Economic Review*, *25*(3), 418-447.

Barrett, C. B., Ghezzi-Kopel, K., Hoddinott, J., Homami, N., Tennant, E., Upton, J., & Wu, T. (2021). A scoping review of the development resilience literature: Theory, methods and evidence. *World Development*, *146*, 105612.

Barrett, C. B. (1996). On price risk and the inverse farm size–productivity relationship. Journal of Development Economics, 51(2), 193–215.

Barrett, C. B., Bellemare, M. F., & Hou, J. Y. (2010). Reconsidering conventional explanations of the inverse productivity–size relationship. *World development*, *38*(1), 88-97.

Barrett, C. B., & Bevis, L. E. (2015). The self-reinforcing feedback between low soil fertility and chronic poverty. *Nature Geoscience*, *8*(12), 907-912.

Beegle, K., Himelein, K., & Ravallion, M. (2012). Frame-of-reference bias in subjective welfare. Journal of Economic Behavior & Organization, 81. https://doi.org/10.1016/j. jebo.2011.07.020.

Beegle, K., C. Carletto, and K. Himelein. 2012. "Reliability of recall in agricultural data." Journal of development economics 98:34–41.

Beg, Sabrin, Mahnaz Islam, and Khandker Wahedur Rahman. "Information and behavior: Evidence from fertilizer quantity recommendations in Bangladesh." *Journal of Development Economics* 166 (2024): 103195.

Benami, E., Z. Jin, M. R. Carter, A Ghosh, Robert J. Hijmans, Andrew Hobbs, Benson Kenduiywo, and David B. Lobell. (2021). "Uniting remote sensing, crop modelling and economics for agricultural risk management." *Nature Reviews Earth & Environment* 2(2): 140-159.

Benjamin, D. (1995). Can unobserved land quality explain the inverse productivity relationship?. Journal of Development Economics, 46(1), 51–84.

Berazneva, J., McBride, L., Sheahan, M., & Güereña, D. (2018). Empirical assessment of subjective and objective soil fertility metrics in east Africa: Implications for researchers and policy makers. *World Development*, *105*, 367-382.

Berazneva, J., Maertens, A., Mhango, W., & Michelson, H. (2023). Paying for agricultural information in Malawi: The role of soil heterogeneity. *Journal of Development Economics*, *165*, 103144.

Bevis, L. E., & Barrett, C. B. (2020). Close to the edge: High productivity at plot peripheries and the inverse size-productivity relationship. *Journal of Development Economics*, *143*, 102377.

Bohr, N., Deisemann, T., Gollin, D., Kosmowski, F., & Lybbert, T. J. (2024). The Seeds of Misallocation: Fertilizer Use and Maize Varietal Misidentification in Ethiopia. *Available at SSRN 4689857*.

Bold T, Kaizzi KC, Svensson J, Yanagizawa-Drott D. (2017). Lemon technologies and adoption: measurement, theory and evidence from agricultural markets in Uganda. *Quarterly Journal of Economics*. 132(3):1055–1100.

Bold, T., Ghisolfi, S., Nsonzi, F., & Svensson, J. (2022). Market access and quality upgrading: Evidence from four field experiments. *American Economic Review*, *112*(8), 2518-2552.

Bulte, E., Di Falco, S., Kassie, M., & Vollenweider, X. (2023). Low-quality seeds, labor supply and economic returns: experimental evidence from Tanzania. *Review of Economics and Statistics*, 1-33.

Burke, M., Lobell, D.B. (2017). Satellite-Based assessment of yield variation and its determinants in smallholder African systems. Proc. Natl. Acad. Sci. 114, 2189–2194.

Burlig, F., Jina, A., Kelley, E. M., Lane, G. V., & Sahai, H. (2024). *Long-range forecasts as climate adaptation: Experimental evidence from developing-country agriculture* (No. w32173). National Bureau of Economic Research.

Cai, Y., K. Guan, D. Lobell, A. B. Potgieter, S. Wang, J. Peng, T. Xu et al. (2019) "Integrating satellite and climate data to predict wheat yield in Australia using machine learning approaches." *Agricultural and forest meteorology* 274: 144-159.

Carletto, Calogero, Sydney Gourlay, and Paul Winters. 2015. "From Guesstimates to GPStimates: Land Area Measurement and Implications for Agricultural Analysis." Journal of African Economies 24, no. 5:593–628. doi:10.1093/jae/ejv011.

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

Carletto, C., Gourlay, S., Murray, S., & Zezza, A. (2017). Cheaper, faster, and more than good enough: is GPS the new gold standard in land area measurement?. In *Survey Research Methods* (Vol. 11, No. 3, pp. 235-265).

Carletto, G., Beegle, K., Himelein, K., Kilic, T., Murray, S., Oseni, M., Scott, K., & Steele, D. (2010). Improving the availability, quality and policy relevance of agricultural data: The Living Standards Measurement Survey—Integrated Surveys on Agriculture. Third Wye City Group Global Conference on Agricultural and Rural Household Statistics. http://www.fao.org/fileadmin/templates/ess/pages/rural/wye_city_group/2010/May/WYE_2010.2.1_Carletto. pdf.

Carter, M., de Janvry, A., Sadoulet, E., & Sarris, A. (2017). Index insurance for developing country agriculture: a reassessment. *Annual Review of Resource Economics*, *9*(1), 421-438.

Chi, G., Fang, H., Chatterjee, S., & Blumenstock, J. E. (2022). Microestimates of wealth for all low-and middle-income countries. *Proceedings of the National Academy of Sciences*, *119*(3), e2113658119.

Cohen, A. (2019). Estimating farm production parameters with measurement error in land area. *Economic Development and Cultural Change*, *68*(1), 305-334.

Corral, C., Giné, X., Mahajan, A., & Seira, E. (2020). Autonomy and specificity in agricultural technology adoption: evidence from Mexico. *NBER working paper*, (w27681).

Dar, M. H., De Janvry, A., Emerick, K., Sadoulet, E., & Wiseman, E. (2024). Private input suppliers as information agents for technology adoption in agriculture. *American Economic Journal: Applied Economics*, *16*(2), 219-248.

Deines, J. M., Patel, R., Liang, S. Z., Dado, W., & Lobell, D. B. (2021). A million kernels of truth: Insights into scalable satellite maize yield mapping and yield gap analysis from an extensive ground dataset in the US Corn Belt. *Remote sensing of environment*, *253*, 112174.

Deininger, K., Jin, S., & Ma, M. (2022). Structural transformation of the agricultural sector in low-and middle-income economies. *Annual Review of Resource Economics*, *14*(1), 221-241.

Deininger, K., Carletto, C., Savastano, S., & Muwonge, J. (2012). Can diaries help in improving agricultural production statistics? Evidence from Uganda. *Journal of Development Economics*, *98*(1), 42-50.

De Weerdt, J., Gibson, J., & Beegle, K. (2020). What can we learn from experimenting with survey methods?. *Annual Review of Resource Economics*, *12*(1), 431-447.

Desiere, S., & Jolliffe, D. (2018). Land productivity and plot size: Is measurement error driving the inverse relationship?. *Journal of Development Economics*, *130*, 84-98.

Dillon, A., Carletto, G., Gourlay, S., Wollburg, P., & Zezza, A. (2021). Agricultural survey design: lessons from the LSMS-ISA and beyond, LSMS Guidebook. *Washington DC: World Bank*.

Djima, I. Y., & Kilic, T. (2024). Attenuating measurement errors in agricultural productivity analysis by combining objective and self-reported survey data. *Journal of Development Economics*, *168*, 103249.

Miguel, J. D. N. (2024). Returns to quality in rural agricultural markets: Evidence from wheat markets in Ethiopia. *Journal of Development Economics*, 103336.

Elbers, C., Lanjouw, J. O., & Lanjouw, P. (2003). Micro-level estimation of poverty and inequality. *Econometrica*, *71*(1), 355-364.

Euler, M., Krishna, V. V., Jaleta, M., & Hodson, D. (2022). Because error has a price: A systematic review of the applications of DNA fingerprinting for crop varietal identification. *Outlook on Agriculture*, *51*(4), 384-393.

FAO, World Bank, & UN Habitat. (2019). Measuring Individuals' Rights to Land: An Integrated Approach to Data Collection for SDG Indicators 1.4.2 and 5.a.1. Washington, DC: FAO, World Bank, and UN Habitat.

Fabregas, R., Kremer, M., Lowes, M., On, R., & Zane, G. (2024). *Digital information provision and behavior change: Lessons from six experiments in East Africa* (No. w32048). National Bureau of Economic Research.

Foster, A. D., & Rosenzweig, M. R. (2022). Are there too many farms in the world? labor market transaction costs, machine capacities, and optimal farm size. *Journal of Political Economy*, *130*(3), 636-680.

Gars, J., Kishore, A., & Ward, P. (2015). Confidence and information usage: Evidence from soil testing in India. *Unpublished Working Paper*.

Gollin, D., & Udry, C. (2021). Heterogeneity, measurement error, and misallocation: Evidence from African agriculture. *Journal of Political Economy*, **129**(1), 1–80.

Gourlay, S., Aynekulu, E., Shepherd, K., & Carletto, C. (2017). Collecting the dirt on soils: Advancements in plot-level soil testing and implications for agricultural statistics. Policy Research Working Paper No. 8057.

Gourlay, S., Kilic, T., Martuscelli, A., Wollburg, P., & Zezza, A. (2021). High-frequency phone surveys on COVID-19: good practices, open questions. *Food policy*, *105*, 102153.

Gourlay, S., Kilic, T., & Lobell, D. (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, 102376. https://doi.org/10.1016/ j.jdeveco.2019.102376.

Haggblade, S., Diarra, A., Jiang, W., Assima, A., Keita, N., Traore, A., & Traore, M. (2021). Fraudulent pesticides in West Africa: a quality assessment of glyphosate products in Mali. *International Journal of Pest Management*, *67*(1), 32-45.

Harou, A.P., Madajewicz, M., Michelson, H., Palm, C.A., Amuri, N., Magomba, C., Semoka, J.M., Tschirhart, K. and Weil, R., 2022. The joint effects of information and financing constraints on technology adoption: Evidence from a field experiment in rural Tanzania. *Journal of Development Economics*, *155*, p.102707.

Henderson, J. V., Storeygard, A., & Weil, D. N. (2012). Measuring economic growth from outer space. *American economic review*, *102*(2), 994-1028.

Hoel, J. B., Michelson, H., Norton, B., & Manyong, V. (2024). Misattribution prevents learning. *American Journal of Agricultural Economics*.

Kanyanda, S., Markhof, Y., Wollburg, P., & Zezza, A. (2021). Acceptance of COVID-19 vaccines in sub-Saharan Africa: evidence from six national phone surveys. *BMJ open*, *11*(12), e055159.

Kern, A., Barcza, Z., Marjanović, H., Árendás, T., Fodor, N., Bónis, P., Bognár, P. and Lichtenberger, J., 2018. Statistical modelling of crop yield in Central Europe using climate data and remote sensing vegetation indices. *Agricultural and forest meteorology*, *260*, pp.300-320.

Kilic, T., Koolwal, G. B., & Moylan, H. G. (2020). Are you being asked? Impacts of respondent selection on measuring employment. *Impacts of Respondent Selection on Measuring Employment (February 18, 2020). World Bank Policy Research Working Paper*, (9152).

Kim, C., E. Lentz, K. Baylis, H. Michelson (2024). "Discordance across household food insecurity indicators: implications for measurement and early warning". *Draft, January 2022.*

Klein, H. A. (1974). *The science of measurement: A historical survey*. Courier Corporation.

Kosmowski, F., Abebe, A., & Ozkan, D. (2020). Challenges and lessons for measuring soil metrics in household surveys. Geoderma. https://doi.org/10.1016/j.geoderma.2020. 114500.

Jean, N., Burke, M., Xie, M., Davis, W. M., Lobell, D. B., & Ermon, S. (2016). Combining satellite imagery and machine learning to predict poverty. *Science*, *353*(6301), 790-794.

Josephson, A., Kilic, T., & Michler, J. D. (2021). Socioeconomic impacts of COVID-19 in low-income countries. *Nature human behaviour*, *5*(5), 557-565.

Lang, Stefan, Petra Füreder, Barbara Riedler, Lorenz Wendt, Andreas Braun, Dirk Tiede, Elisabeth Schoepfer et al. "Earth observation tools and services to increase the effectiveness of humanitarian assistance." *European Journal of Remote Sensing* 53, no. sup2 (2020): 67-85.

Lobell, D., Tommaso, S., You, C., Djima, I., Burke, M., & Kilic, T. (2019). Sight for Sorghums: Comparisons of satellite- and ground-based Sorghum yield estimates in Mali. Remote Sensing, 12, 100.

Lobell, D. B., Azzari, G., Burke, M., Gourlay, S., Jin, Z., Kilic, T., & Murray, S. (2020). Eyes in the sky, boots on the ground: Assessing satellite-and ground-based approaches to crop yield measurement and analysis. *American Journal of Agricultural Economics*, *102*(1), 202-219.

Malapit, H., Quisumbing, A., Meinzen-Dick, R., Seymour, G., Martinez, E. M., Heckert, J., E.M., Heckert, J., Rubin, D., Vaz, A., Yount, K.M. (2019). Development of the project-level Women's Empowerment in Agriculture Index (pro-WEAI). *World development*, *122*, 675-692.

Maxwell, D., Vaitla, B., & Coates, J. (2014). How do indicators of household food insecurity measure up? An empirical comparison from Ethiopia. *Food policy*, *47*, 107-116.

Marenya, P.P., and C.B. Barrett. 2009. State-Conditional Fertilizer Yield Response on Western Kenyan Farms. American Journal of Agricultural Economics 91 (4): 991–1006.

Michelson, H., Gourlay, S., Lybbert, T., & Wollburg, P. (2023). Purchased agricultural input quality and small farms. *Food Policy*, *116*, 102424.

Michelson, H., Fairbairn, A., Ellison, B., Maertens, A., & Manyong, V. (2021). Misperceived quality: fertilizer in Tanzania. *Journal of Development Economics*, *148*, 102579.

Michler, J. D., Josephson, A., Kilic, T., & Murray, S. (2022). Privacy protection, measurement error, and the integration of remote sensing and socioeconomic survey data. *Journal of Development Economics*, *158*, 102927.

Naugler, A., H. Michelson, S. Janzen (2024). Agri-dealer Entry and Exit in Tanzania: Implications for Markets and Consumers. *Working paper, January 2024*.

Panek, E., & Gozdowski, D. (2020). Analysis of relationship between cereal yield and NDVI for selected regions of Central Europe based on MODIS satellite data. *Remote Sensing Applications: Society and Environment*, *17*, 100286.

Parris, T. M., & Kates, R. W. (2003). Characterizing and measuring sustainable development. *Annual Review of environment and resources*, *28*(1), 559-586.

Poets, A., Silverstein, K., Pardey, P. G., Hearne, S., & Stevenson, J. (2020). DNA fingerprinting for crop varietal identification: Fit-for-purpose protocols, their costs and analytical Implications.

Rios, A. R., & Shively, G. E. (2005). Farm size and nonparametric efficiency measurements for coffee farms in Vietnam. In Paper presented at the American agricultural economics association annual meeting, Providence, RI.

Rosenzweig, M. R., & Udry, C. (2014). Rainfall forecasts, weather, and wages over the agricultural production cycle. *American Economic Review*, *104*(5), 278-283.

Rudin-Rush, L., Michler, J. D., Josephson, A., & Bloem, J. R. (2022). Food insecurity during the first year of the COVID-19 pandemic in four African countries. *Food Policy*, *111*, 102306.

Saenger, C., Torero, M., and Qaim, M. (2014). Impact of third-party contract enforce- ment in agricultural markets—a field experiment in Vietnam. *American Journal of Agricultural Economics*, 96(4):1220–1238.

Sanabria, J., Ariga, J., Fugice, J., & Mose, D. (2018). Fertilizer quality assessment in markets of Uganda. USAID and IFDC.

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

Schwalbert, R. A., Amado, T., Corassa, G., Pott, L. P., Prasad, P. V., & Ciampitti, I. A. (2020). Satellite-based soybean yield forecast: Integrating machine learning and weather data for improving crop yield prediction in southern Brazil. *Agricultural and Forest Meteorology*, *284*, 107886.

Steele, Jessica E., Pål Roe Sundsøy, Carla Pezzulo, Victor A. Alegana, Tomas J. Bird, Joshua Blumenstock, Johannes Bjelland et al. "Mapping poverty using mobile phone and satellite data." *Journal of The Royal Society Interface* 14, no. 127 (2017): 20160690.

Quinn, J. A., Nyhan, M. M., Navarro, C., Coluccia, D., Bromley, L., & Luengo-Oroz, M. (2018). Humanitarian applications of machine learning with remote-sensing data: review and case study in refugee settlement mapping. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *376*(2128), 20170363.

Rabbi, I.Y., Kulakow, P.A., Manu-Aduening, J.A., Dankyi, A.A., Asibuo, J.Y., Parkes, E.Y., Maredia, M.K., 2015. Tracking crop varieties using genotyping-by-sequencing markers: a case study using cassava (Manihot esculenta Crantz). BMC Genet. 16 (1), 1–11.

Stevenson, J., Macours, K., & Gollin, D. (2018). The rigor revolution in impact assessment: Implications for CGIAR.

Tamim, A., Harou, A., Burke, M., Lobell, D., Madajewicz, M., Magomba, C., Michelson, H., Palm, C. and Xue, J., 2022. Relaxing Credit and Information Constraints: Five-Year Experimental Evidence from Tanzanian Agriculture.

Tjernstrom, E. (2016). *Signals, similarity and seeds: Social learning in the presence of imperfect information and heterogeneity* (No. B157). FERDI Policy Brief.

Upton, J., Constenla-Villoslada, S., & Barrett, C. B. (2022). Caveat utilitor: A comparative assessment of resilience measurement approaches. *Journal of Development Economics*, *157*, 102873.

Van der Weide, R., Blankespoor, B., Elbers, C., & Lanjouw, P. (2022). How accurate is a poverty map based on remote sensing data?. *Policy Research Working Paper*, *10171*.

Varshney, K. R., Chen, G. H., Abelson, B., Nowocin, K., Sakhrani, V., Xu, L., & Spatocco, B. L. (2015). Targeting villages for rural development using satellite image analysis. *Big Data*, *3*(1), 41-53.

Weiss, M., Jacob, F., & Duveiller, G. (2020). Remote sensing for agricultural applications: A meta-review. *Remote sensing of environment*, *236*, 111402.

Wheeler, T., von Braun, J., 2013. Climate change impacts on global food security. Science 341, 508.

Wineman, A., Anderson, C. L., Reynolds, T. W., & Biscaye, P. (2019). Methods of crop yield measurement on multi-cropped plots: Examples from Tanzania. *Food security*, *11*(6), 1257-1273.

Wollburg, P., Tiberti, M., & Zezza, A. (2021). Recall length and measurement error in agricultural surveys. *Food Policy*, *100*, 102003.

Wollburg, P., Bentze, T., Lu, Y., Udry, C., & Gollin, D. (2024). Crop yields fail to rise in smallholder farming systems in sub-Saharan Africa. *Proceedings of the National Academy of Sciences*, *121*(21), e2312519121.

Wooldridge, Jeffrey. 2008. Introductory Econometrics: A Modern Approach. Mason, OH: Cengage Learning.

Wossen, T., Abdoulaye, T., Alene, A., Nguimkeu, P., Feleke, S., Rabbi, I.Y., Manyong, V., (2019), Estimating the productivity impacts of technology adoption in the presence of misclassification. *American Journal of Agricultural Economics*. 101 (1), 1–16.

Yeh, C., Perez, A., Driscoll, A., Azzari, G., Tang, Z., Lobell, D., Ermon, S. and Burke, M., 2020. Using publicly available satellite imagery and deep learning to understand economic well-being in Africa. *Nature communications*, *11*(1), p.2583.

Yigezu, Y. A., Alwang, J., Rahman, W., Mollah, M. B., El-Shater, T., Aw-Hassan, A., et al. (2018). Is DNA fingerprinting the gold standard for estimation of adoption and impacts of improved lentil varieties? Food Policy, 83.