

Agricultural Data Collection to Minimize Measurement Error and Maximize Coverage

Calogero Carletto

Andrew Dillon

Alberto Zezza



WORLD BANK GROUP

Development Economics
Development Data Group
July 2021

Abstract

Advances in agricultural data production provide ever-increasing opportunities for pushing the research frontier in agricultural economics and designing better agricultural policy. As new technologies present opportunities to create new and integrated data sources, researchers face trade-offs in survey design that may reduce measurement error or increase coverage. This paper first reviews the econometric and survey methodology literatures that focus on the

sources of measurement error and coverage bias in agricultural data collection. Second, it provides examples of how agricultural data structure affects testable empirical models. Finally, it reviews the challenges and opportunities offered by technological innovation to meet old and new data demands and address key empirical questions, focusing on the scalable data innovations of greatest potential impact for empirical methods and research.

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

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

Agricultural Data Collection
to Minimize Measurement Error and Maximize Coverage

Calogero Carletto
World Bank

Andrew Dillon
Northwestern University

Alberto Zezza
World Bank

Keywords: Agriculture, Measurement Error, Sampling Error, Survey Design, Data Collection

Acknowledgments: We are fortunate to have close collaborators who have greatly influenced our thinking on agricultural data collection including past and present colleagues at the World Bank Living Standards Measurement Study team, national statistical office collaborators, Innovations for Poverty Action, and the Global Poverty Research Lab. We are grateful for comments on the draft from two anonymous reviewers, Leah Bevis, Sarah Kopper, Karen Macours, Christopher Udry, and the Handbook's editors Christopher Barrett and David Just. We appreciate research support from Raka Banerjee. This working paper is a pre-typeset version of a chapter prepared for the Handbook of Agricultural Economics, Volume 5: Agricultural Production and Research Methods, edited by Chris Barrett and David Just.

Table of Contents

1. Introduction.....	4
2. Minimizing Measurement Error	8
2.1. <i>Questionnaire design</i>	<i>13</i>
2.2. <i>Interviewer effects.....</i>	<i>16</i>
2.3. <i>Respondent effects.....</i>	<i>17</i>
2.4. <i>Mode of data collection.....</i>	<i>20</i>
2.5. <i>Processing errors.....</i>	<i>21</i>
3. Trade-offs in Maximizing Coverage.....	22
3.1. <i>Sampling frame.....</i>	<i>24</i>
3.2. <i>Units of analysis.....</i>	<i>26</i>
3.3. <i>Survey timing.....</i>	<i>30</i>
3.4. <i>Mode of data collection.....</i>	<i>31</i>
3.5. <i>Attrition.....</i>	<i>33</i>
4. Empirical Specification, Data Structure, and Measurement Error	34
4.1. <i>Profit and production functions</i>	<i>34</i>
4.2. <i>The agricultural household model</i>	<i>38</i>
5. Advances in Data Collection	40
5.1. <i>Advances in selected thematic areas.....</i>	<i>41</i>
5.2. <i>Advances in data collection modes and data structures</i>	<i>60</i>
6. Conclusions.....	68
Bibliography	72

1. Introduction

In the past two decades, innovations in data systems have led to the production of more real-time, disaggregated, and interoperable data on agriculture than ever before. Increasing data demands and emerging policy questions are driving much of this innovation, with fast technological change and methodological advances providing an opportunity to collect more and better data at lower costs (Akogun et al., 2020; Carletto et al., 2015; Dillon et al., 2021a; Kosmowski et al., 2019; Liao, 2018; Lobell et al., 2019). Investments in country-level data infrastructure have enabled new approaches to methodological innovation, such as incorporating randomized control trials into national panel data collection or devising improved methods to ensure greater data interoperability. Meanwhile, new types of data – such as remote sensing data and citizen-generated data – and new technologies – such as portable sensors, DNA fingerprinting, and computer-assisted personal interviewing (CAPI) – provide unparalleled prospects for collecting and analyzing a wide array of agricultural constructs in a more granular, timely, and cost-effective manner. These advantages are further enhanced by integrating new types of data with traditional data sources such as household surveys, censuses, and administrative data.

While other data sources are becoming increasingly important, household and farm surveys are likely to remain the centerpiece of policy research for agricultural and development economists. Not only are household surveys a key data source in their own right, but they serve as interoperable complements and validation instruments for other data sources, such as for the ground-truthing of remote sensing data, or for the ex-post adjustment of bias in studies based on citizen-generated and other non-probability data. Emerging literature on a wide array of agricultural measurement issues in land, production, and gender analysis has relied upon innovations in survey design, as fostered in the past decade through data initiatives like the Living Standards Measurement Study-Integrated Surveys on Agriculture (LSMS-ISA) and the Global Strategy to Improve Agricultural and Rural Statistics (GSARS).

The influential publication on household survey data collection by Grosh and Glewwe (2000), and in particular the chapter by Reardon and Glewwe (2000) on agriculture, together with other chapters on consumption, income, and enterprises, provided an original contribution to the field of

survey measurement issues that remains relevant to this day, as does the influential work by Sudman and Bradburn (1974) on response effects in the United States. However, significant innovations in methodological development for household surveys have taken place in recent years, including on the collection of agricultural data in multi-purpose surveys. Agricultural survey design continues to evolve through important innovations such as scaling up the collection of plot-level data in low-income countries, gender-disaggregated agricultural data,¹ agricultural panel surveys, and the collection of national agricultural household and enterprise data,² *inter alia*.

While the importance of household and farm surveys within national agricultural data systems is indisputable, it is equally important to recognize their limitations in addressing new data challenges. For instance, household and farm surveys may be ill-suited to capture the evolving value chains of rapidly transforming agri-food systems (Barrett et al., forthcoming). Surveys seldom collect sufficient data on contracting and on the different agents involved in transactions with the household and, when they do, they tend to be case studies focused on a few commodities in limited geographies or be qualitative in nature³ (Barrett et al., 2020; Minten et al., 2016).

Furthermore, surveys often lack sufficient spatial and temporal resolution and are unable to provide the real-time data needed by policy makers, being limited by cost and sample size considerations. In higher income countries, remote sensing has been widely used for decades as a complement to ground-based measures for an array of applications, including sample frame construction, crop area and land use estimation, crop conditions assessment, climate data, and production forecasting (Hale et al. 1999). In recent years, the use of Earth Observation data for agricultural applications, combined and validated with ground-based measurements, has been spreading rapidly in low- and middle-income countries, yielding promise for more accurate and timely agricultural data in these contexts (Lobell et al., 2019; Gourlay et al., 2019).

¹ See Doss and Quisumbing (this volume citation) for an extensive review of gender-disaggregated data.

² Agricultural sector censuses such as FAO's World Programme for the Census of Agriculture includes agricultural households and agricultural enterprises. For a recent review of this program, see WCA (2020).

³ An exception in national surveys is the collection of network data in a few LSMS-ISA surveys, where information is collected from respondents on agents involved in the transaction of agricultural inputs and outputs.

Unfortunately, despite impressive progress in both traditional and new data sources, large gaps still persist in terms of the availability and quality of agricultural data. Furthermore, mounting global challenges such as rising inequality, climate change, and rapid population growth remain are likely to disproportionately affect the agriculture sector and rural areas, with more significant impacts for low- and middle-income countries. Meanwhile, the ongoing COVID-19 pandemic provided a stark reminder of the need to accelerate the production of more timely and accurate data to save lives. The pandemic has also exposed growing inequities in data systems across countries, with innovation moving at a faster pace in higher-income countries (United Nations and World Bank, 2020). Worse still, agricultural data gaps tend to be the largest where good data are needed the most, that is, in resource-constrained countries for which agriculture represents the lifeline of the majority of households and the whole economy. At the same time, the emergence and diffusion of complex farms in higher income countries (Kling and Mackie, 2019; Macdonald, 2016) creates new layers of difficulty in data collection and measurement. Recognizing that individual data sources are often unable to singlehandedly address these complex and multifaceted challenges, researchers are increasingly focusing on the potential offered by improved data integration and interoperability between data sources.

While appreciating the importance of improving agricultural data in all countries along the entire income gradient, this paper intentionally focuses on some of the data challenges and scalable applications and tools most suitable to low- and middle-income countries. Because of this geographic focus, we primarily limit our discussion to household and farm surveys, as they are likely to remain the instrument of choice and backbone of agricultural data systems in many countries for years to come. The attention to surveys is also warranted by the availability of a fully developed total survey quality framework around which we develop the narrative of the paper. The growing attention to survey design issues and a burgeoning literature on rigorous survey methodological experiments (de Weerd et al., 2020) also provide added motivation for the focus of the paper.

In this paper, we will argue and provide evidence that renewed attention to data quality issues – specifically in terms of measurement error and data coverage – is critical for advancing the research frontier in agricultural economics and designing better agricultural policy. Both

measurement error and issues of limited data coverage threaten the internal and external validity of empirical analysis on agriculture, constraining its efficacy and relevance in informing sectoral policies and investments. A better understanding of measurement error and error-generating processes is crucial, as errors negatively affect the accuracy and validity of inferences resulting from data, and thus limit the usefulness of data to policy making. Given the significance of these issues, agricultural economists and survey practitioners have paid increasing attention to measurement error in recent years, drawing on insights from existing literature on labor economics, survey methodology, and statistics. The fact that this is the first paper is fully dedicated to measurement and data is testament to the prominence that data, in general, and measurement issues, in particular, have acquired in the profession today. The purpose of this paper is to demonstrate that improving agricultural data structures – that is, making agricultural data systems more credible and fit-for-purpose – can address both measurement error and coverage issues to facilitate better empirical analysis on agriculture. For our purposes, we define data structure as the full set of survey design choices that comprise the data production process, including sampling, questionnaire design, and fieldwork implementation.

Today, technology and a well-piloted modernization agenda offer the opportunity to push the data quality production frontier, both in terms of availability and quality of data. Furthermore, increasing demands for evidence-based policy making and accountability have generated the tailwind to achieve critical advances in agricultural data in general, and agricultural survey data in particular. Addressing existing flaws in survey data would greatly contribute to raising the credibility and, ultimately, the quality of the resulting research and analysis (Jerven and Johnston, 2015). Achieving the “credibility revolution” in empirical research as advocated by Angrist and Pischke (2010) calls for better research design choices, which begins with addressing measurement error and coverage issues. Making agricultural research more policy-relevant, credible, and fit-for-purpose begins with improving the quality of its underlying data to expand the set of testable empirical models.

This paper highlights the importance of improving agricultural data structures for empirical analysis, while accounting for the inherent trade-offs intrinsic to designing data collection for agricultural research and policy analysis. In the section that follows, we review sources of

measurement error from the perspective of the economics, survey methodology, and statistics literatures, referring to this rich bibliography for a more detailed discussion of the issues. In section three, we turn to design choices related to coverage, including sampling design, the unit of analysis, survey timing, data collection modes, and attrition. The fourth section integrates sources of measurement error and coverage biases to assess their implications and trade-offs in the empirical specification of a few examples of agricultural models, documenting where innovation in data structure has advanced the research frontier. The fifth section offers innovative approaches for addressing measurement error and coverage biases in agricultural data, based on recent technological advances and foreseen opportunities. In the sixth and final section, we conclude with recommendations on priorities for accelerating improvements in the accuracy and coverage of agricultural data, ultimately to support higher-quality research for better agricultural policy.

2. Minimizing Measurement Error

Measurement error and related issues of non-random measurement error have been discussed in some of the earliest work by Fisher (1926) and Working (1925). Since then, these topics have been extensively articulated and well-documented across many subdisciplines in economics, such as health, labor, industrial organization, and applied welfare analysis (Bound et al., 2001; Chesher and Schluter, 2002; De Haan et al., 2019; Gottschalk and Huynh, 2010; Hu and Schennach, 2008; Hyslop and Imbens, 2001; Pischke, 1995; Schennach, 2016, 2004; Rom et al., 2020). Most of these papers consistently highlight that bias induced in parameter estimates depends on the structure of the measurement error found in the data, as well as the identifying assumptions that empirical economists make when estimating those parameters. Making the right assumptions for these structures and tackling the sources of errors, at both the design and analytical stages, can greatly improve the accuracy and relevance of agricultural data.

While the field of statistics boasts a rich and longstanding literature on measurement error (Biemer, 2010, 2009; Biemer et al., 1991; Biemer and Lyberg, 2003; Carroll et al., 2006; Deming, 1944; Groves, 1989; Groves and Lyberg, 2010; Kasprzyk, 2005; Kish, 1965; Wansbeek and Meijer, 2000), we have only more recently witnessed a burgeoning literature in agricultural and development economics journals addressing the sources, magnitude, and implications of

measurement error, and proposing new ways to validate and correct for measurement error biases. Measurement error can result in both bias and variable error, or variance. With non-random measurement error biases in parameter estimation come faulty conclusions and misguided policies. Even with random measurement error, increased statistical noise requires larger sample sizes to identify parameters of interest, increasing the cost of data collection. Hence, we again emphasize the importance of understanding the sources of measurement error and attenuating its impact.

In the field of survey methodology, the Total Survey Error (TSE) framework has been the dominant paradigm. The framework serves as a useful organizing structure for assessing the extent and composition of different sources of errors that affect estimates, guiding researchers and data collection practitioners towards appropriate design choices for minimizing measurement error and maximizing coverage (Groves and Lyberg, 2010). TSE “refers to the accumulation of all errors that may arise in the design, collection, processing and analysis of survey data” (Biemer, 2010). The paradigm implies that total errors must be minimized for a given budget and that the major sources of errors should be identified and prioritized to achieve maximum accuracy for a given cost (Biemer, 2010). Broadly speaking, TSE can be viewed as encompassing the concept of data quality which, in statistical terms, is partially captured by the Mean Square Error (MSE), a metric of the accuracy of the estimated variable.

Minimizing measurement error in agricultural data has been problematic due to a number of inherent features of agricultural processes, particularly for certain crops and agronomic practices in smallholder farming. These features include the highly seasonal nature of production and the irregularity of inputs required in the sequencing of production. Multiple studies have shown that across a variety of issues, farmers’ self-reported information, which often involves long recall periods, has proven to be inadequate (Beegle et al., 2013; Deininger et al., 2011; Fermont and Benson, 2011; Gourlay et al., 2017).

Although long aware of the existence of measurement error, only recently have agricultural economists shown interest in how these errors affect their inferences and the policy recommendations deriving from their analysis. Even when measurement errors were considered, the common practice was to make rather cavalier suppositions about the property and distribution

of the errors by assuming a classical measurement error (CME) – that is, assuming that the error in the variable of interest is independent from its true value as well as from the measurement errors in all other variables in the model and the stochastic error term. While reliance on the CME assumption can be justified in some instances, it is seldom the case for many variables, for which the error-generating process appear to follow more complex and systematic patterns that fail the classical assumption. The assumption appears to be even more troublesome for non-linear models (Bound et al., 2001). More recently, the agricultural economics literature has aptly focused on the potential systematic biases resulting from measurement error and how design choices and new technologies can help improve measurement (for some recent applications of non-classical measurement error in agricultural data, see Abay et al., 2019; Carletto et al., 2013; Desiere and Jolliffe, 2018; Gourlay et al., 2017). We argue that addressing potential bias ex-ante through appropriate design choices may ultimately be a more effective way to tackle the issue, although careful ex-post analysis and modeling may also be helpful in mitigating its impact on estimates (Gollin and Udry, 2021; Maue et al., 2020).

Policy researchers hold the power and responsibility to make wiser design choices at the data collection stage for given objectives and budget constraints. To this end, the TSE framework provides a useful blueprint for understanding the underlying error-generating processes and the relative importance of the different components, as well as how to ameliorate their impact on estimates. While the TSE framework is useful for this paper, given its focus on sample surveys as one of the main sources of data for policy research in agriculture, it is important to note that most features of TSE also apply to other data sources. For instance, Biemer (2017) argues that TSE provides very useful insights on how to deal with errors in Big Data, drawing clear parallels between errors in surveys and the often selective, incomplete, and erroneous nature of Big Data-generating processes. As researchers increasingly rely on alternative data sources such as citizen-generated data and crowdsourcing to collect agricultural data, similar data quality frameworks should be developed for those types of data. However, even in the case of TSE, full consensus on a comprehensive typology of errors is yet to exist. Groves and Lyberg (2010) conclude that this lack of consensus is the natural consequence of the continuous evolution of methods and data collection technologies, as well as the different objectives and constraints of different data

producers and analysts. As a result, any list defining the universe of TSE is bound to be incomplete and/or to emphasize certain components over others (Groves and Lyberg, 2010).

We must note here that focusing solely on minimizing total survey error with expensive measurement methods ignores the research design cost-variance trade-off and the full set of research design choices. For instance, a researcher may be willing to accept some degree of measurement error, if reducing such error would also reduce the statistical power of the research design. If a researcher is implementing a randomized control trial, measurement error that is not correlated with treatment status may not bias estimates, whereas in a non-experimental design, measurement error might bias parameter estimates and thus have consequences for internal validity and policy recommendations.

To conceptualize these research design trade-offs more clearly, Dillon et al. (2020) build on earlier writing in the statistical literature (Biemer, 2010, among others) to introduce the idea of the *data quality production function*. For any given research project, the researcher's objective is to maximize the knowledge or evidence generated from the research project. To do so, the researcher makes decisions about the identification strategy, statistical power, and external validity of the project, subject to a budget constraint and the data quality production function. The data quality production function includes choices on questionnaire design as well as other variables such as sampling, empirical approach, and field implementation modes, protocols, and constraints. These latter choices include decisions based on the availability of financial resources, personnel capacity, and the competing demands and/or mandates of the researcher or agency collecting the data.

Thus, measurement error and bias, which closely relate to the concept of internal validity, must be weighed against other important features of model inferences, including the power of the estimates, external validity and coverage, and the intended use of the data (Dillon et al., 2020). From a user's perspective, data accuracy (and the costs involved in achieving it) must be weighed against other idiosyncratic user preferences related to the broader construct of *fitness-for-use* of the data (Juran and Gryna, 1980) as part of a broader Total Survey Quality (TSQ) framework (Biemer, 2010). This more complete construct of survey quality, going beyond accuracy, includes concepts such as comparability, relevance, timeliness, accessibility, credibility, usability, interpretability,

completeness, and coherence (Biemer, 2010). For instance, the temporal or spatial granularity of the estimates and other features related to improved coverage may be more important to some users, who may be willing to sacrifice some degree of accuracy in exchange. Another highly relevant dimension is the interoperability of the data and how data integration can improve accuracy and decrease bias while also playing a role in enhancing and/or reducing coverage. For instance, the use of mixed-mode data collection – such as high-frequency phone surveys that are fully integrated into a less-frequent face-to-face large-scale survey – has the potential to reduce measurement errors due to recall bias, but may introduce other problems such as under-coverage due to the incompleteness of sampling frames or higher levels of attrition. As proposed by Biemer and Lyberg (2003), one could treat all these additional dimensions of quality as constraints in an error minimization problem (Biemer, 2010). While highly relevant for sample surveys, the total survey quality paradigm can also be extended to other sources of data (Amaya et al., 2020).

Keeping in mind the specific design choices that researchers face, we define the possible sources of measurement errors – corresponding to what Groves (1989) calls *errors of observation* – into five groupings: (1) questionnaire design, (2) interviewer effects, (3) respondent effects, (4) mode of data collection, and (5) data processing. Equally important sources of errors may derive from incomplete coverage, or lack of representativeness (that is, *errors of non-observation*), including sampling errors as well as non-sampling errors, further categorized into coverage errors and non-response – we address these in the following section.

This taxonomy of sources of errors can be juxtaposed with a typical data structure – with *units of observation* in the rows and *variables* in the columns – to show the relationship, and thus potential trade-offs, between sources of measurement errors affecting *variables* (the columns) vis-à-vis non-coverage errors affecting *units of observation* (the rows), as well as the trade-offs between measurement error and coverage. It must be noted, however, that many of these sources and design choices may affect both measurement error and coverage (e.g., mode effects) or be correlated and have covariate effects on total error (e.g., interviewer and respondent effects). Furthermore, sources of measurement errors are likely to simultaneously affect multiple variables, both dependent and independent, generating complex error structures that have differential implications on inferences. Hausmann (2001) reviews approaches to dealing with measurement error in either

dependent or independent variables and in the case of continuous and discrete variables. Hyslop and Imbens (2001) provide a clear and succinct classification of the effect of measurement error on either dependent or independent variables, as well as on both. A common approach to measurement error in empirical labor economics is to model measurement error as an ‘errors in variables’ problem whose proposed solution is an instrumental variable. However, increased concerns about weak instruments have caused such methods to be disfavored in labor economics and this approach to measurement error in empirical agricultural economics has been rare. Finally, continuous dependent variables may lead to reduced statistical precision but not necessarily bias – but the cost of increasing sample size (adding more rows), particularly for numerous sample strata and domains of inference, is often high.

While econometric approaches are inherently ex-post solutions to measurement error that take the data-generating process as given, we see opportunities for ex-ante solutions within current international efforts to build capacity in data quality assurance and methodological innovation in national statistical offices. These capacity building initiatives provide an opportunity to create better agricultural data structures that address research hypotheses and policy concerns by maximizing data quality. To this end, with due consideration to the various trade-offs, researchers can make design choices in several areas to minimize measurement error in the collection of agricultural data. Below, we present in detail the five main sources of measurement errors listed above. Understanding these groupings and their potential impact on bias and variance can help researchers make the right design choices for their research objectives.

2.1. Questionnaire design

Agricultural questionnaire design requires researchers and policy makers to clearly outline the unit of analysis and agricultural processes that they would like to measure. Rozelle (1991) outlined various approaches to agricultural survey design, such as production function approaches, income state approaches, and balance sheet approaches, each of which require different questionnaire designs. Production function approaches map inputs to outputs to estimate the returns to inputs. Income statement designs measure farm profits based on revenue and expense information. A balance sheet approach values farm assets and liabilities in addition to inputs and outputs. An early

resource for agricultural questionnaire design is the Reardon and Glewwe (2000) agricultural chapter in Ghosh and Glewwe (2000), which outlines broad principles of agricultural module design in multi-topic household surveys. Dillon et al. (2021a) provide a recent updated review incorporating recent innovations in survey design choices for agricultural questionnaires, including the integration of plot-level crop production and input modules as well as livestock production questionnaires.

A broad questionnaire design literature explores best practices to minimize measurement error. Errors from questionnaire design may result from unclear wording, poor formatting, priming, excessive length of questions and instrument, sequencing and skipping of questions, duration of reference period, and differences in reference periods or the coding of responses (Schwarz, 1997; Fowler, 1995; Gideon, 2012; Iarossi, 2006; Manski and Molinari, 2008; Payne, 1980; Sudman and Bradburn, 1973; Sudman et al., 1996; de Weerd et al., 2020). The impact of questionnaire design choices on data quality can be substantial, with even minor changes having adverse consequences on the accuracy and comparability of estimates (Beegle et al., 2020; Das et al., 2012; De Weerd et al., 2016). Specification errors, which occur “when the concept implied by the survey question and the concept that should be measured in the survey differ” (Biemer, 2010) can also contribute to errors from poorly designed questionnaires. One example of specification error in many agricultural surveys is lack of clarity when defining plots relative to parcels, which may have large implications for productivity estimates (see section 3.2 on units of analysis for a discussion of plots versus parcels). Lack of consistent specification in the definition of household membership or contextual differences in the social and economic criteria of household membership may also lead to faulty estimates (Beaman and Dillon, 2012). Other examples of questionnaire design choices are the use of rosters to collect individual or plot-level data, or the collection of individual components of income or profits in lieu of eliciting information in a more aggregated format (De Mel et al., 2008; Vijverberg and Mead, 2000). In both paper and CAPI-based questionnaires, visual aids are widely used for capturing non-standard units of measurement for more accurate estimations of both agricultural production and food consumption (Eisenhower et al., 1991; Mathiowetz, 2000; Oseni et al., 2017). Particularly in electronic questionnaires, area maps using GPS are increasingly used for estimating land area, for listing dwellings and plots for sampling, as well as for supervision and quality control purposes.

Questionnaire length and complexity, as well sequencing of the questions and modules, may also have important implications on measurement error (Kilic and Sohnesen, 2019; Strack et al., 1988; Schuman and Presser, 1981; Schwarz and Hippler, 1991). Furthermore, the use of open or closed format may have an impact on responses, where closed questions with clearly identified response options can help respondents in both remembering information and choosing appropriate responses (Kasprzyk, 2005; Schwarz and Hippler, 1991). Finally, the language(s) and translation of the questionnaire, as well as differences in language and cultural background between the survey designer and the respondent may also contribute to errors (Vaessen et al, 1987).

Agricultural surveys have attempted to reduce measurement error by leveraging the number of visits within the agricultural season to reduce the length of recall and align the visits to key stages of production. There are obvious costs to this approach, with a higher number of visits possibly increasing respondent and interviewer fatigue as well as field costs. Meanwhile, the potential advantages including reducing the length of the recall period, breaking up the length of long, complex questionnaires and interviews, or using the first interview to identify respondents for specific follow-ups in a second interview or as a temporal reference point to help respondents better contextualize their answer (known as bounding questions). Evidence from the measurement of consumption clearly points to an excessive number of visits negatively affecting data quality, most likely linked both to respondent fatigue and interviewer effect (Engle-Stone et al., 2017 for nutrient consumption in a survey employing up to 7 visits over 14 days in Bangladesh; Schündeln, 2018 for a consumption survey with households visited up to 10 times in Ghana).

Several of the surveys supported by the LSMS-ISA program (Dillon et al., 2021a) have attempted to adjust the survey visit schedule to the calendar of the agricultural campaign by scheduling a post-planting and a post-harvest visit. This visit structure aimed to ease the cognitive burden of the respondents by asking questions on agricultural operations and harvest at most a few weeks or months after they occur, instead of 12 months or more, while also limiting the number of visits to contain survey costs and respondent burden. Using data for Tanzania and Malawi, Wollburg et al. (2020) confirm the presence of non-random measurement error systematically related to the length of the recall period. They find evidence of such error in all the main variables of interest in any

agricultural survey, including quantities harvested, labor and fertilizer inputs, and even the number of cultivated plots. The magnitude of the recall effect typically varies between two and five percent per additional month of recall length, rendering its impact on the reliability of key agricultural indicators economically significant. Recently, some African national statistical offices such as the Uganda Bureau of Statistics are experimenting with an additional visit that can be used to collect supplementary objective measures on farm plots, including crop cuts (Ponzini et al., 2021). Technology is aiding these innovations by facilitating the transfer of information across survey visits via increasingly flexible CAPI applications.

2.2. Interviewer effects

Interviewer effects occur when personal characteristics of the interviewer, such as education, ability, motivation, or language barriers, affect the interview process. Proper recruitment, training, and monitoring of job performance are used to minimize errors associated with interviewer effects (Fowler, 2004). A meta-analysis of the literature (West and Blom, 2017) establishes that interviewer behavioral traits and demographic characteristics influence survey responses, and by extension, data quality. Response rates and response biases are particularly influenced by specific interviewer characteristics (such as age, ethnicity, experience, and education), behaviors (such as formal versus conversational interview styles), cognitive and non-cognitive skills (such as mathematical ability, reading, attention to detail, and empathy), and interviewer experience. Existing literature on interviewer effects finds that data quality can be a function of who is asking the questions. Responses vary by the interviewer's ethnicity (Davis et al, 2010; Davis and Silver, 2003), gender (Benstead, 2010; Flores and Lawson, 2008) and religion (Blaydes and Gillum, 2013), especially for questions sensitive to race, gender, and religion respectively. Studies have also explored the association of data quality with interviewer skills and behaviors such as probing, providing feedback for responses, and rapport building (Belli et al, 2004). Some interviewer characteristics are fixed, while skill-based characteristics may change in response to training.

Responses and measurement error may also vary based on the interviewer's adherence to a script. For instance, in the context of the Agricultural Labor Survey in the United States, Ridolfo et al. (2021) show how interviewers' lack of adherence to the script resulted in significant measurement errors. Similarly, using the same survey, Rodhouse et al. (2019) quantify the extent to which

deviating from the script affects the likelihood of measurement errors and conclude that the presence of measurement error is highly associated with the interviewer's ability to adhere to the script. Biagas et al. (2019), using the same data, use a novel multi-method approach to identify patterns of interviewer behavior and its contribution to total survey error.

Recent research on interviewer effects in a randomized experiment in Uganda by Di Maio and Fiala (2019) found that interviewer characteristics and their differences from respondent characteristics affected survey responses and ultimately data quality for sensitive topics. On the contrary, responses to less sensitive topics were much less, or not at all, susceptible to interviewer characteristics. This is supported by additional research suggesting that the salience and sensitivity of the questions influence the nature and magnitude of interviewer effects (Himelein, 2015; Laajaj and Macours, 2021). Marx et al. (2018) provide evidence on the impacts of team composition and ethnic diversity on interviewer performance. Data on the time use of field teams suggests that teams composed solely of interviewers organize tasks more efficiently than teams that include supervisors, interviewers, and data monitors, which demonstrate lower levels of effort. In a review of several studies, Groves (1989) suggests that demographic traits result in interviewer effects only when the specific question is related to the demographic characteristics of the interviewer (i.e., an interviewer effect based on the race of the interviewer may be found for questions about race). This may be particularly relevant in contexts with large ethnic and racial diversity.

The effect of priming in surveys and the inconsistent application of interviewing protocols and wording across interviewers is also likely to generate systematic biases (Lavrakas, 2008). Similarly, the interview setting may also affect interviewers' recording of responses and result in systematic errors. Collecting detailed metadata on the interview process is often used to partially control for potential biases generated by poor interview settings; unfortunately, this practice is not consistently applied across surveys.

2.3. Respondent effects

Respondents can also contribute to TSE in several additional ways, either intentionally or unintentionally. Assumptions about the structure of those respondent biases are often uninformed by empirical evidence, although Hyslop and Imbens (2001) provide a categorization of different

types of potential biases. For instance, respondents may intentionally under-report the amount of land they own because of taxation concerns or may conversely over-report their land holdings because of prestige considerations or social desirability. Social desirability concerns are likely to result in the under-reporting of “socially undesirable” behavior, and the over-reporting of socially desirable occurrences (Bound et al., 2001). For some agricultural statistics such as child labor, context may determine whether children’s work in agriculture carries social stigma and hence potential reporting bias. Similar response behavior may also occur with the reporting of income variables (Tourangeau et al., 2000). Respondents may also round up the amount of land owned to integer values, resulting in a phenomenon known as heaping. Research on land area measurement consistently finds systematic errors in farmers’ self-reporting, with farmers that own smaller land holdings systematically over-reporting (and farmers with larger land holdings under-reporting), as well as considerable heaping in reporting (Carletto et al., 2015, 2013).

Errors in responses may be unintentional, resulting from limited knowledge or recall bias due to memory decay as the length of the recall period increases. Errors may also derive from limited understanding of the questions, potentially correlated with the cognitive level of the respondent (Laajaj et al., 2019; Laajaj and Macours, 2021). The use of bounding techniques, providing an easy-to-remember temporal reference point in respondents’ memory to better contextualize the answer, can be used to reduce the effect of telescoping in recall (Abate et al., 2020; Neter and Waksberg, 1964). Recall biases are also affected by the salience of the event being recalled (Beegle et al., 2012; Gaddis et al., 2020; Kilic et al., 2021; Wollburg et al., 2020). Gaddis et al. (2020) and Arthi et al. (2017) analyze the impact of recall on the measurement of agricultural labor. Their findings suggest that a seasonal recall approach to agricultural labor measurement may result in underestimated labor productivity. In their cross-country study, Beegle et al. (2012) find no evidence of bias in harvested quantities for both staple and cash crops. Recall bias, however, was present in hired labor reporting, although the direction of these biases varied by country. As already mentioned, similar findings emerge from Wollburg et al. (2020) related to the design choices of number and timing of field visits, as more visits and shorter roll-out periods reduce the length of the recall period.

Interestingly, at least in domains outside of agriculture, perceptions of salience may be influenced by the length of the recall period (Winkielman et al., 1998) and may vary by the income level of the respondent (Das et al., 2012). Understanding respondents' cognitive strategies is crucial for choosing the most appropriate length of recall and thus minimizing measurement errors in responses. Evidence suggest that beyond a certain recall length, respondents switch from enumeration to estimation, each translating into different errors (de Nicola and Giné, 2014; Scott and Amenuvegbe, 1991).

The use of proxy respondents and widespread reliance on the most informed respondent – often the male head of the household – is also likely to result in biased responses (Dillon and Mensah, 2020; Doss et al., 2019; Kilic et al., 2020; Kilic and Moylan, 2016; Krosnick, 1999; Moore, 1988). Bardasi et al. (2011) find that using proxy responses led to the under-reporting of men's participation rates in agricultural activities. Kilic et al. (2020a) show significant impacts of respondent selection strategy in the collection of labor data. Kilic et al. (2020b) also find that the common practice of proxy reporting results in different reporting of land assets relative to those reported by self-respondents. The use of proxy reporting by the “most knowledgeable household member” results in higher rates of exclusive ownership of agricultural land among men, and lower rates of joint ownership among women, as compared to the gold standard approach of individual, self-respondent interviews (Kilic et al., 2020b). In this context, interview setting has also been shown to greatly affect responses. For instance, the common practice of non-private interviewing (i.e. where other members of the household and community may be present during the interview), more often conducted through proxies, results in significant under-reporting of employment relative to measurement through private, self-respondent interviews, with stronger effects for women than men (Kilic et al., 2020a). Dillon and Mensah (2020) note that when proxies report household-level agricultural variables as opposed to individual-level responses, proxy response bias is composed of both aggregation errors and asymmetric information within the household. Thus, their findings suggest that proxy response bias is not solely due to asymmetric information within the household, as is commonly assumed in the literature on proxy response bias for individual-level variables.

Measurement error can also derive from the use of peers (e.g. neighbors, co-workers, key informants, etc.) as proxy respondents, potentially resulting from projection or false consensus biases, among other things (Hogset and Barrett, 2010). Despite the potential biases of proxy reporting of peer behaviors (Ashenfelter and Krueger, 1994), the practice is widely used (Hogset and Barrett, 2010). In some cases, however, it may be justified when gathering data from peers may be sub-optimal yet preferable, such as in the case of collecting highly sensitive information. While the use of proxy respondent should be minimized to the extent possible, one must also acknowledge the trade-offs between respondent bias and coverage, as restricting interviewing to the selected respondents is likely to result in higher attrition and unit missingness. Furthermore, the use of proxy respondents is often unavoidable due to logistics or cost considerations. In such cases, the proxy respondent selection process should be conducted based on strict standardized field protocols.

2.4. Mode of data collection

The mode of data collection – whether face-to-face, self-administered or interviewer-led, by phone, or by web, and whether on paper or in electronic format – can have substantial effects on measurement error as well as coverage. In terms of measurement error, several studies have shown that the effect depends on the type of question as well as interviewer ability and respondent characteristics (Biemer and Lyberg, 2003; Caeyers et al., 2010; De Leeuw, 2005; De Leeuw and Van der Zouwen, 1988). While errors of coverage may result from incompleteness of frames and respondent selection, phone or web surveys may allow more frequent data collection for greater temporal granularity and lower measurement error due to recall bias. Similarly, crowdsourcing and other forms of citizen-generated data are increasingly used in agriculture and potentially offer enormous opportunities for collecting data at greater temporal and spatial resolution. However, these modes of data collection also exhibit serious limitations in terms of representativeness as well as overall data quality which, if left unaddressed, are bound to produce biased inferences (Aceves-Bueno et al., 2017; Japac et al., 2015; Wiggins et al., 2011; Ambel et al., 2021; Brubaker et al., 2021). Using data from several African countries, Brubaker et al. (2021) address the issue of representativeness of phone surveys for gender-level estimates based on individual non-random respondents in the household, most commonly the head of the household, and propose ways to mitigate the bias.

Gibson et al. (2017) summarize the literature on data quality comparing phone to face-to-face interviewing, much of which focuses on health rather than agricultural variables. A significant increase in methodological work related to data quality on phone surveys occurred in response to the COVID-19 pandemic. Das et al. (2021) and Dillon et al. (2021b, c) provide summaries of random digit dialing phone surveys related to improving response rates, optimal timing of phone survey call attempts, and the impact of pre-survey text messaging.

Greenleaf et al. (2020) estimates a difference of 20 percentage points in reported contraceptive use, with higher reported use for phone interviewing. In this case, presumably because sensitive subjects may induce lower reporting, phone interview modes may decrease measurement error by providing respondents more anonymity. In Lamanna et al. (2018), phone interviewing methods also induced higher reporting in dietary measures, with increases that varied between 28 percentage points, 14 percentage points, and 18 percentage points for minimum dietary diversity, minimum meal frequency, and minimum acceptable diet measures, respectively. Mahfoud et al. (2014) compared estimates of alcohol consumption and exercise, finding a four percentage point increase in alcohol consumption and a seven percentage point decrease in exercise level reported with phone survey mode relative to face-to-face interviewing. However, not all studies find statistically significant impacts on data quality by survey mode; for example, Gallup (2012) finds no differences between survey modes in an experiment in Honduras. Furthermore, much of the existing evidence does not test behavioral mechanisms that might explain why responses differ by survey mode (Tourangeau and Yan 2007).

2.5. Processing errors

Finally, processing errors include possible errors generated during data entry, editing, coding, weighting, and analysis of data. New technologies and data processing power have transformed the set of opportunities and ways to reduce processing errors. Unfortunately, the relatively faster growth in the volume of data being generated, combined with the complexity of the new data landscape, have created additional challenges in terms of processing errors. For a review of processing errors and how they may impact total error, see Biemer and Lyberg (2003).

Many authors emphasize and empirically demonstrate how new data sources, methods, and technologies have proven instrumental to attenuating the sources of measurement error in data collection, as discussed in section 5 of this paper (Abay et al., 2019a, 2020; Dillon et al. 2019; Gourlay et al., 2019; Kosmowski et al., 2019; Lobell et al., 2020). In particular, the recent proliferation of well-designed validation studies, relying on complementary data sources and readily available technology, is contributing to a better understanding of the relationship between the measured and true value of the variable of interest, as well as providing insights on the magnitude and direction of potential bias. For instance, recent advances in CAPI, the use of sensors and other direct measurements, and the use of metadata and paradata, are increasingly being used to offset the threats to data quality of more traditional data collection techniques that rely heavily on farmers' self-reporting (Akogun et al., 2020; Pratt et al., 2020; Sinha et al., 2020).

3. Trade-offs in Maximizing Coverage

As described in the previous section, measurement error can be addressed ex-ante through design choices, or ex-post through proper analysis and modeling, particularly when error-generating processes are non-random, thus potentially resulting in biased inferences. Aside from measurement error due to the poor design and administration of survey instruments, there are several other non-sampling errors which negatively affect researchers' analyses if left unaddressed. Of particular concern are *coverage errors*, which occur when individuals or units of interest are excluded from the sample, resulting in serious repercussions for the validity of the estimates, as they fail to represent the entire population or area of interest. Most troublingly, errors of coverage are seldom random and tend to exclude subpopulations of interest, such as female farmers, smallholders living in remote areas, pastoralist and nomadic populations, and more distant plots, among other marginal groups. They may also result from the use of particular sampling frames such as population-based listings (based on the population living in the area covered by the survey) or area frames (based on geographic area covered in the survey) which, by construction, exclude or undercount certain subgroups. For example, using population-based listings to collect agricultural data for the estimation of total national agricultural production or the farm size-productivity relationship prevents the inclusion of larger commercial farms in the list. The exclusion of commercial farming

from population-based listings may also hamper the analysis of the potential spillovers in terms of labor and other inputs for neighboring smallholders. For instance, Ali et al. (2019), using data from Ethiopia, find little or no benefits in terms of job creation, input market access, or technology transfer due to the presence of large farms. The possible exclusion or undercounting of medium- and large-scale farms from population-based listings may also constrain analysis on the rapid process of agricultural transformation and land consolidation occurring in many countries (Jayne et al. 2019). The advantages of population-based sampling frames relative to area frames have also been questioned, based on the assertion that respondents tend to under-report on plot ownership and use, thus leading to the underestimation of total production. On the other hand, the use of area frames may result in systematic biases in the collection of socioeconomic variables from the plot owners. The use of multiple frame sampling is often advocated as a way to overcome the limitations of either approach (Gonzalez Villalobos and Wigton, 2011; FAO, 2015b).

Coverage errors can also derive from omitted variables in model specification due, for instance, to missing environmental variables affecting production decisions (Sherlund et al., 2002). As seldom collected in surveys, information on environmental conditions and capturing inter-farm heterogeneities affecting farmers' choice is often missing, potentially resulting in biased estimates of both estimates and coefficients. With the widespread availability of inexpensive geospatial data that can be linked to household-level data, filling these gaps and potentially reducing the risk of omitted variable bias has become increasingly easier. However, the georeferencing of survey data at farm and plot level is yet to become common practice in many low- and middle-income countries. Furthermore, many of the hurdles related to data privacy for secure and confidential dissemination and use remain unresolved.

In addition to non-coverage errors due to a priori exclusion from the frame, non-response and other reasons for attrition are also likely to affect the validity of the estimates. Non-response can be further divided into unit and item non-response. Unit non-response occurs when a unit of concern is included in the frame but is either not reached or is unwilling to participate in the survey. This missingness is seldom random and is often treated ex-post through imputation methods (Kilic et al., 2020; Rubin, 1996, 1987). Conversely, item non-response occurs when a respondent fails to provide information to a question during interviewing. This missingness is most often handled at

the data processing stage through complex imputation methods. However, the lack of consistent guidelines and practices of value imputation, combined with poor documentation on how missing values have been treated – potentially leading to systematic errors – once again highlights the potential trade-offs between coverage and measurement error.

In line with the broader total survey quality framework, we review the issue of coverage error found in the statistical literature, considering several of the dimensions of coverage related to *fitness-for-use*. As postulated above, researchers are likely to be faced with trade-offs between measurement error and other dimensions of data quality. Furthermore, these trade-offs are also related to the particular data source. For instance, the feasibility of various possible decisions on the unit of analysis, timing, or level of disaggregation will vary based on whether data is being collected through an agricultural census, a national survey, or a randomized control trial. Coverage error is also affected by the mode of data collection, such as face-to-face, phone, or web-based interviewing, which can also lead to mode effects, another factor contributing to total error. In this section, we describe some of the design choices related to sampling frames, units of analysis, survey timing, modes of data collection, and attrition, all of which have implications for data coverage and the trade-offs between measurement error and coverage.

3.1. Sampling frame

The sampling frame used for agricultural surveys, whether based on population listing or geographic areas, may often be missing some sub-populations or units of concern. Under-coverage could be accidental or intentional and, if deliberate, may be driven by many motives. Of particular relevance to smallholder agriculture is the issue of under-coverage of more remote areas, farm holdings, or individual plots, driven by either cost considerations or convenience (Kilic et al., 2017). Capturing pastoralist and transient populations is also particularly challenging, and their exclusion is likely to result in biased estimates (Himelein et al., 2014).

As stated above, coverage will also depend on the type of sampling frame chosen. Multi-purpose household surveys like the LSMS-ISA use a population-based listing, with the household as the unit of analysis. More recent agricultural surveys, particularly in more developed economies, have

increasingly relied on area or point frames, which presents advantages when the objective is to estimate production at the national or sub-national level. In fact, one concern with using population-based listings is the possibility of missing some plots due to misreporting, thus resulting in lower total production. This problem is further compounded for pastoralist and semi-nomadic populations which are entirely missing from or hard to reach using population frames; in these cases, using area frames in combination with population listings may be more appropriate (Himelein et al., 2014). However, collecting socioeconomic information on the farm household starting from a given area or point is particularly challenging, and hardly ever done in socioeconomic studies, thus limiting the analytical use of the data. Finding ways of reconciling the choice of frames and maximizing coverage by linking multiple frames has been the subject of recent research under the 50x2030 Data Smart Agriculture initiative (D’Orazio, 2020).

One key shortcoming of using a population frame for agricultural data collection is the wholesale exclusion of medium and large commercial farms. This has been the practice in agricultural household surveys such as those conducted under the LSMS-ISA initiative, which has raised concerns about the validity of inferences made on a truncated distribution of farm holdings (Muyanga and Jayne, 2019; Ali and Deininger, 2014). On the other hand, agricultural censuses and farm surveys that focus on both farming households and commercial farms may be more suitable for sector-level estimation of agricultural indicators, but fall short of meeting the analytical objectives of surveys of households as both production and consumption units (Singh et al., 1986). The use of multi-frame sampling strategies, combining the strengths of the individual sources, is gaining ground in lower-income countries, albeit at a slower pace due to capacity constraints and the technical difficulties involved.

By choosing a household listing as a population-based frame, the definition of household and household membership have important measurement implications, in part because social and economic definitions of the household diverge (Beaman and Dillon, 2012). This is especially the case in communities with extended farming families with land inheritance claims, common production of family lands, or complicated land use rights. The household definition matters from an agricultural perspective, as household membership defines the individuals that will be included in modules on agricultural land holdings, labor, assets, and marketing decisions. Unsurprisingly,

reported values of agricultural production are lower when the household definition excludes some agricultural producers. Residency requirements also complicate the measurement of pastoralist activities when households are involved in transhumant pastoralism.

3.2. Units of analysis

Agricultural data is often collected at different units of analysis, including at national or sub-national levels, geospatial area units, and at the household, farm, plot, or plot-crop-season-manager levels. Depending on the frame used, selection may be based on population listing or a map, subdivided into grids. It may be the case that, for a specific application, both types of frames may be used, which requires reconciling the estimates at the national or sub-national level. For instance, Pelletier et al. (2020) use small area estimation (SAE) to reconcile deforestation estimates from area-based frames with smallholders' use of modern inputs drawn from a population-based frame. In fact, it is conceptually appealing to use area frames when the focus is the measurement of land area and related agronomic features, as a population-based survey may result in under-counting plots and farming activities. On the other hand, the use of population-based listings may be more appropriate than area frame sampling in capturing data from hard-to-reach farmers living far away from the plots. The use of multiple frames, which in the agricultural domain involves the combined use of both area and list frames, has been advocated as a way to ensure completeness of the frame and full coverage of the sector (FAO, 2015b; Gonzalez Villalobos and Wigton, 2011). However, avoiding duplication and overlapping units is often a challenge when constructing multi-frames.

Indirect sampling has also been proposed as a way to overcome the shortcomings intrinsic to household listings as frames for agricultural statistics, by applying Generalized Weighted Sampling Methods (GWSM) to obtain estimates for landholdings (the unknown and more relevant universe) from a household listing (the known population) (Falorsi et al., 2016; Gennari et al., 2013).

Any livestock sector outcomes should be recorded at the herd level rather than at the household level, particularly for nomadic populations. Household surveys with population-based sampling

frames will never capture herd-level outcomes, for which area-based-sampling measures may be a better unit of analysis.

In an example of innovative survey design, LSMS-ISA surveys redesigned agricultural modules by recognizing that the unit of observation for agricultural production is often not the household but the plot, which may be managed by differing household members, with significant implications for sex-disaggregated analysis. In early versions of LSMS surveys and many other multi-topic household surveys, agricultural activities were not detailed at the plot level, as this requires a higher respondent burden relative to household level recall. However, a plot-level approach more accurately measures the relationship between inputs and outputs in agricultural production. Production heterogeneity results from differences in crops cultivated that require different levels of inputs and may not be managed by the household head. The level of detail required in subsequent modules makes this survey design choice non-trivial. For example, plot-level data collection requires not simply measuring land area at the plot level, but also production, labor, capital, chemical inputs, and land management techniques.

Choosing to measure production at the plot level requires a wider series of choices in identifying the unit of analysis, which has implications for both design and implementation. First, agricultural production is seasonal and multiple crop cycles on a given plot need to be measured. Second, plots are not always associated with a single crop, as multi-cropping or inter-cropping is a common land management practice for increasing yield and land quality. In contexts such as smallholder agriculture in Sub-Saharan Africa, inter-cropping is the norm, not the exception. For instance, two-thirds of maize plots in Uganda include multiple crops (authors' calculations based on the 2018 Uganda National Panel Survey). Third, property rights are not necessarily well established in many rural contexts, and multiple family members may work on the plot or cultivate the plot in different seasons. The landowner may be different than the person making decisions to cultivate the land, as sharecropping, land leasing, or land lending may mean that landowners are not making agricultural decisions. Holden et al. (2016) describe survey design choices in modules used to describe land tenure. Depending on the empirical applications of data collected, information on the sources of land acquisition including inheritance and legal status, land transactions, formal and informal property rights, land conflicts, perceptions of tenure security, and trust in land-related

institutions may all be important additional questions complementary to the land roster. This is especially true if the survey aims to monitor the Sustainable Development Goals (SDGs) related to land tenure, particularly SDGs 5.a.1 and 1.4.2. Recognizing the importance of land tenure security as it relates to the control of and access to other assets, through access to credit and investment in land, for example, these SDG indicators seek to measure specific aspects of land tenure at the individual level, rather than at the household level.

In summary, the key innovation in conducting plot-level production analysis is not to simply measure inputs and outputs at the plot level, but to distinguish the unit of analysis as plot-crop-season-manager. This unit of analysis facilitates comprehensive measurement of household production, allowing multiple analytical strategies from seasonal, crop, and gender perspectives, but also has some limitations, particularly in the context of a panel survey, given the changing demarcation of plots across seasons. Tracking parcels over time is often a more feasible option.

Across different agricultural systems, the vocabulary associated with an agricultural landholding may also differ. Farmers use different words to indicate their farms, parcels, and plots, often with contradictory meanings. It is important that any agricultural survey design reflects a clear conception of the hierarchy of units consistent with the agricultural system that is being measured. Carletto et al. (2016) provide an overview of land area measurement survey design issues, noting differences in units of land measurement as well as variation across LSMS-ISA surveys in land reporting units. Holdings, parcels, fields, and plots all have internationally accepted definitions, although their interpretations by both academics, NSOs, and policy makers often lead to ambiguity. The definition of the agricultural holding is the primary unit of analysis in agricultural surveys, whereas the household is the primary unit of analysis in household surveys. The Food and Agriculture Organization of the United Nations (FAO) (2015) defines the agricultural holding as an “economic unit of agricultural production under single management comprising all livestock kept and all land used wholly or partly for agricultural production purposes, without regard to title, legal form or size...” Holdings can be divided into parcels, and the FAO notes that “a distinction should be made between a parcel, a field and a plot”, where “a field is a piece of land in a parcel separated from the rest of the parcel by easily recognizable demarcation lines such as paths, cadastral boundaries, fences, waterways or hedges. A field may consist of one or more plots, where

a plot is a part or whole of a field on which a specific crop or crop mixture is cultivated, or which is fallow or waiting to be planted” (FAO 2015). However, when designing and implementing an agricultural survey, practitioners should confirm the tiers and definitions used by the national statistical office, as these may not always coincide with the FAO definitions. As suggested above, tracking parcels may be a more practical option for longitudinal studies, given the changing size of plots across seasons.

Recording agricultural information at the household level inherently aggregates individual production and imposes a linearity assumption across plots for input utilization and asset use. The main trade-off in recording agricultural information at the plot level is that farmers must recall input allocation at the plot level, which requires more cognitive effort and response time. These recall biases may be compounded by proxy response bias, as plot-level self-reporting is time-consuming in the field and may be not feasible for all survey responses. Proxy respondents may have incomplete information on plots managed by other household members. For farmers who purchase inputs collectively with their family for multiple plots, it may be difficult to accurately assess how much fertilizer, seed, or other input was applied to a particular plot. Consistent with time use data, it may also be difficult for a farmer to recall individual household labor allocations to particular plots over an agricultural season or with respect to particular agricultural tasks. While more research is needed to understand the measurement implications of the disaggregation of input and production data to the plot level from the household level, the known analytical advantages of doing so – such as analysis of male-managed plots vis-à-vis female-managed plots – outweigh the unknown risk of aggregation in many surveys, including LSMS-ISA surveys.

Due to variation in land tenure status and land use rights, it is also important to account for seasonality in production on plots and changes in plot management when considering the unit of analysis. Depending on the agricultural season, a plot may be cultivated by a different member of the household and use different levels of inputs along with different cropping choices. Researchers have often cited asymmetries in crop type and input use, and therefore productivity and earnings, by the gender of the plot manager. O’Sullivan et al. (2014) estimate that, after controlling for plot size and region, productivity differences across male- and female-managed plots in Africa ranged from 23 to 66 percent. In order to appropriately account for plot-level production, and to enable

analysis of the timing of production and gender asymmetries, both season and plot manager should be considered. The plot manager may differ from one season to the next, often depending on gender-based norms.

Just as there are trade-offs in empirical specifications among units of analysis, differences in units of analysis also imply different constraints when repeated observations are an objective of the survey design.

3.3. Survey timing

Survey timing is a critical design choice that affects coverage as well as measurement error due to questionnaire design or respondent effects. Survey timing can refer to the timing of visits within a single agricultural season, as well as the timing of visits between seasons. LSMS-ISA surveys feature an innovative survey design that includes repeated visits both within and between seasons (Carletto et al., 2010). New international surveys such as those produced under the 50x2030 Initiative will also feature multiple within- and between-season observations. Here, we review survey timing decisions that increase coverage across certain dimensions of time.

As a research and policy issue, seasonality and the presence of multiple cropping cycles within the agricultural calendar imply that multiple agricultural surveys may be best timed according to cropping cycles or the agricultural calendar. Survey timing choices can decrease recall bias, particularly for agricultural choices such as input decisions, labor allocation, or sales, which are often frequent and difficult to recall. LSMS-ISA surveys collect agricultural information during the post-planting and post-harvest periods, but timing could be more frequent to correspond to multiple cropping periods which may overlap, particularly when agricultural systems have fewer water constraints. When variables such as labor inputs are a research objective, higher frequency surveys may reduce recall bias as well as increase statistical power depending on the auto-correlation of the variable over periods of time (McKenzie 2012).

Agricultural panel surveys most frequently track households between seasons to capture changes in production decisions over time, and their correspondence to changes in welfare. Despite interest

in understanding changes in agricultural activities over time, plots are rarely tracked in panel surveys, as tracking plots over time is a time-intensive field activity that limits coverage. For example, LSMS-ISA surveys are conducted as a household panel (or household-parcel panel in select countries), with repeated cross-sections of a tracked household's plot, production, and input information.

Variation in production and other agricultural variables as well as the ability to monitor shocks and household resilience could also be captured through community sentinel sites complementing less frequent surveys (Barrett and Headey, 2014). The authors convincingly argue for establishing a multi-country system of sentinel sites in selected communities as a way to improve the timeliness and coverage of agricultural data, in the face of ever more frequent shocks affecting the resilience of rural households.

3.4. Mode of data collection

As discussed in the previous section, the choice of survey mode may have significant implications in terms of measurement error, either directly or through its interaction with other design choices related to questionnaire design, interviewer selection, and respondent features. Similarly, certain modes of data collection may also affect survey coverage. Poor representativeness because of inadequacy of the sample frame, selectivity, and potentially high attrition is a major challenge for phone surveys (Ballivian et al., 2015; Gibson et al., 2019). For instance, the use of mobile phones or the web affects not just how responses are elicited, but also whether respondents agree to participate in the survey, and/or whether respondents are included in the frame in the first place. In terms of frames, phone surveys predominantly rely on three options: (1) recent representative surveys with phone numbers of respondents; (2) lists of phone numbers from mobile phone providers, and (3) random digit dialing (Kastelic et al., 2020; McGee et al., 2020). Each option involves significantly different implications for both coverage and attrition. Proper tracking and field protocols, combined with the collection of selected information for ex-post weighing and bias adjustment, can greatly enhance the representativeness and usability of phone surveys and reduce mode bias.

Irrespective of the type of frame used, phone surveys are more likely to miss more remote and poorly connected households, as well as poorer households who do not own a phone or live in areas with poor mobile coverage. This is particularly relevant for agricultural data, where large shares of respondents live in remote and poorly connected areas and are more likely to be poor and technologically illiterate. Educational level, age, and technological literacy will also systematically affect overall coverage. Similarly, selection bias for respondents in citizen-generated data and crowdsourcing makes collecting agricultural data using those modes particularly concerning in terms of representativeness and coverage, especially when it comes to their use in official statistics. These concerns, combined with the huge opportunities that these new modes of data collection provide, are generating significant attention in recent literature (see Hill et al., 2019 and some of the papers cited therein, including Buil-Gil et al, 2020; Diego-Rosell et al., 2020; Salganik, 2017).

Coverage biases related to phone survey modes may be due to either non-response or non-completion. Little literature exists on response rates in phone surveys in low-income countries, but recent studies due to COVID-19 restrictions on face-to-face interviewing are generating new evidence. Dillon et al. (2021b) conducted random digit dialing surveys in nine countries, demonstrating variation in response rates as high as 60 percent and as low as 7 percent. They find most coverage biases are due to non-response when respondents do not answer their telephone. Hence, a key survey design feature in telephone surveys is decreasing non-response. Dillon et al. (2021c) evaluated choices on pre-contact notification via text message and time-of-day/day-of-week as potential telephone survey protocol decisions that might reduce non-response. They found that pre-survey text messages did little to improve non-response but did increase survey completion. In the Philippines, pre-survey text messages actually increased non-response, while having no effect in the Colombia, Mexico, or Rwanda samples. In those countries, pre-survey text messaging increased survey completion by between one and four percentage points. In nine countries, time-of-day and day-of-week effects were estimated, with midday interviews increasing participation and evening calls reducing participation (Das et al. 2021). These effects were relatively small, with an effect size of four and eight percentage points over the base pickup and completion rates, respectively. The day-of-week effects varied substantially between countries, with few generalizable day-of-week effects between countries. Within countries, effect sizes were substantial, but often in different direction by country. More country-specific evidence may need

to be generated to reduce non-response, underscoring the importance of understanding local contexts and the differences in time distributions between work and leisure within a country.

The impact of mode of data collection on coverage extended to other survey design features. For instance, the use of diaries, while possibly more accurate than recall modes when properly implemented, may lead to greater under-coverage among illiterate respondents as well as higher non-response among richer households that face higher opportunity costs for answering lengthy diaries. Also, differences in record-keeping across groups of respondents, such as between smallholders and larger-scale farmers, may result in systematic variations based on the chosen method (Lyberg and Kasprzyk, 2004; Silberstein and Scott, 2004). In the next section, we discuss in detail how measurement error and coverage bias affect the empirical estimation of common agricultural models.

3.5. Attrition

Significant coverage biases due to attrition affect both the internal and external validity of empirical work for both randomized control trials and observational panels (Beegle et al., 2011; Falaris, 2003; Outes-Leon and Dercon, 2008; Rosenzweig, 2003; Thomas et al., 2012, 2001; Zabel, 1998). Millan and Macours (2019) discuss attrition in the context of randomized control trials where tracking protocols may affect intent-to-treat effects. In a 10-year panel from Nicaragua, they find that excluding attrited individuals led to an overestimate of their intent-to-treat effect by 35 percent.

Tracking the same respondent over time is challenging in large national surveys. Thomas et al. (2012) and Witoelar (2011) discuss minimizing attrition and improving tracking in the context of a large-scale national survey. The integration of mobile phones and household geo-referenced data increases the traceability of households, but also raises concerns about privacy and data protection. While multiple papers suggest ex-post methods of dealing with attrition (see for example, DiNardo et al., 2006; Millan and Macours, 2019; Wooldridge, 2002), little consensus on ex-ante methods of reducing attrition have been covered in the literature. One notable exception is Olsen (2005), who discusses design features relevant to attrition reduction in the National Longitudinal Survey

of Youth. Thomas et al. (2012) also discuss planning for attrition and protocols for reducing attrition. They find that success in tracking movers depended not only on observable characteristics of respondents, but also the characteristics of interviewers who initially interviewed respondents. Reducing coverage bias due to attrition is likely to be most successful not simply when surveys are designed to track respondents who have moved, but also when initial interviews collect tracking data and interviewers are trained to establish connections with survey respondents.

4. Empirical Specification, Data Structure, and Measurement Error

For any empirical analysis, the set of theoretical models that can be tested is defined by the available data. Each data set has its own data structure, which we defined above as the full set of survey design choices that comprise the data production process, including sampling, questionnaire design, and fieldwork implementation choices. National production surveys such as agricultural censuses imply a specific subset of production models that can be tested. Household surveys that integrate agricultural data, such as LSMS-ISA surveys, are implicitly informed by producer models or agricultural household models, but measurement error or coverage bias can reduce the precision and utility of estimates and restrict the set of testable models. In this section, we review trade-offs in the empirical specification of agricultural models and data requirements. We discuss how survey design choices that increase data coverage present a trade-off in potentially increasing measurement error in prominent empirical models.

While we cannot review the interaction of data structure and empirical specification across all prominent models in agricultural economics given the focus of this paper, it is illustrative to choose a few common specifications to demonstrate how innovations in data structure have expanded the set of testable models. For this purpose, we review examples from the profit and production function literatures and the agricultural household model literature. The estimation and improvements in estimating these models over the last few decades as international household surveys have emerged are directly cited as motivation in Ghosh and Glewwe (2001) among others.

4.1. Profit and production functions

A large literature examines models of the producer problem (Chambers, 1988; Chambers and Quiggin, 2000; Mundlak, 2001) and the specification of the agricultural production function (Pope and Just, 2001). Pope and Just (2003) provide a summary of production technologies and their functional forms. In this earlier production literature, measurement error and coverage bias were central concerns in the field. Pope and Just (2003) specifically discuss coverage bias and its effect on production function specification, as well as the modeling of measurement error. Aggregated district or national data sources led to misattribution of the returns to inputs, as the unit of analysis in the data was not at the producer level where profit-maximizing decisions were made in the theoretical model.

Measurement error due to unobservable decision variables is also a source of bias in production function estimation, but distinguishing measurement error from unobserved heterogeneity and potential misallocation is challenging. Yields can be biased by output or land size measurement as noted by Abay et al. (2020). Inputs such as fertilizer or labor can be biased by errors in both quantity and quality over the relevant recall period. In the case of livestock production, inputs such as medical care and feeding practices may be difficult to attribute within herds. Measurement error in these input and output variables is likely correlated with unobserved heterogeneity in farmer ability. As agricultural production is also characterized by stochastic disturbances such as weather shocks which require similar modeling assumptions to address unobserved farmer heterogeneity, error terms capture multiple sources of stochastic shocks. In principle, researchers can model such errors in the producer problem depending on the data structure.

Pope and Just (2003) consider the case where measurement error found in demand and supply functions is uncorrelated between input and output – i.e., $Y = f(X^*, K, v)$, as opposed to $Y = f(X^* + \delta, K, v)$ where input disturbances affect output directly. The latter case is called errors in optimization or misallocation, where disturbances are interpreted as errors in decision making. Pope and Just (2003) distinguish misallocation from additive measurement error, their errors in variables case, and the errors in uncontrolled conditions case, when disturbances are modeled as errors that affect production after producer decisions are made. For the producer problem, $\pi(I, K) = \max_X \{E_I[Pf(X, K, \varepsilon) - RX]\}$, the supply and demand equations are $X = X^*(I, K) + \delta + \varsigma$ and $Y = f(X^* + \delta, K, \varepsilon) + v$ where δ is misallocation, ς is the errors in measurement of inputs

that do not affect outputs, ε represents stochastic production shocks such as weather, and v represents measurement of outputs. Using agricultural data from the United States, Pope and Just (2003) estimate the model, finding no evidence for measurement error, but cannot reject misallocation.

In a similar spirit, Gollin and Udry (2021) model measurement error, unobserved heterogeneity, and misallocation using panel data from Tanzania and Uganda. Several important identification challenges are addressed due to data structure, in particular, a unit of analysis at the farmer-crop-plot-season level. After explaining differences in production across farms due to observable differences, Gollin and Udry (2021) note that unobserved variation could be due to unobserved land characteristics, risk, measurement error, or misallocation. With repeated panel data of farmers over time, a production function, whose error term is disaggregated among these different unobserved components, can be estimated. Gollin and Udry's (2021) estimates suggest that measurement error and heterogeneity explain two-thirds to three-quarters of productivity differences, while misallocation affects productivity only modestly. In considering the advances in the identification of measurement error and misallocation in the production function, we note from a data structure perspective the trade-off between improved empirical specification of misallocation and measurement error due to survey design. Recall of input allocations at the farmer-crop-plot-season level is ideal, as it permits researchers to map inputs to outputs, but measurement error may actually be increased if farmers do not recall inputs at the farmer-crop-plot-season level. For example, farmers may make bulk fertilizer purchases within their household that are then divided within the household and across the farmer's plots. Precisely recalling the amount of fertilizer applied to a farmer's maize field relative to their inter-cropped legumes may be impossible, even if the farmer knows exactly how much fertilizer was purchased in total.

We note that the Gollin and Udry (2021) identification strategy capitalizes on a farmer panel to disentangle the effects of measurement error from misallocation and farmer unobservables, but we also note that assumptions about the production technology, as in Pope and Just (2003), are required. Estimates of misallocation using different production functions would certainly vary, along with the level of measurement error estimated for each. This provides an important example of the trade-offs between data structure and empirical modeling. Advances in farmer panels allow

Gollin and Udry (2021) to estimate misallocation and measurement error addressing farmer unobservables through farmer fixed effects in their production function.

Coverage biases in profit and production functions also largely depend on the unit of analysis. In national surveys, units of analysis for agricultural data include the household (Reardon and Glewwe, 2000), the agricultural holding (FAO, 2016), or the plot (Carletto et al., 2016). When land is recorded at the household level, aggregation bias and asymmetric information among household farmers may cause landholdings to be misreported (Dillon and Mensah, 2020). Coverage biases could be significant when commercial farms or large farms are omitted using household sampling frames (Muyanga and Jayne, 2019). The improved coverage of plot-level information may also increase measurement error in input reporting, as noted above. High-frequency agricultural surveys are rare and as such, much knowledge about seasonality is due to post-harvest recall rather than within a production season. Models of sequential agricultural decision-making such as Fafchamps (1993) are few.

Foster and Rosenzweig (2010) discuss coverage and measurement error biases in the estimation of profit functions, particularly in the inference of returns to inputs in technology adoption problems. Cross-sectional inference of returns to fertilizer are biased, as farm heterogeneity and low price variation make it difficult to disentangle the marginal product of fertilizer and land quality in explaining differences in profits. Panel data with increased coverage of farmers across seasons do not necessarily improve the identification of returns to inputs even if measurement error is low, because multiple sources of unobserved heterogeneity remain as correlated biases in the production function residual, such as plot unobservables, between-season misallocation, or climatic variation which may be unobservable ex-post.

While coverage biases are a significant constraint in improving estimates of agricultural profit functions, measurement error remains a significant bias, primarily due to labor recall, input quality, and estimating the use of and return to agricultural assets. Akogun et al. (2020), Carletto et al. (2012), and Dillon et al. (2020) document the challenges of agricultural labor recall in household surveys. Not only is plot-level detail of person-days challenging for respondents to recall, but agricultural wages are difficult to accurately measure in household surveys, since much

agricultural labor is household labor. Physical activity measurement of agricultural labor for particularly physical tasks, may be one approach to better measuring labor quality or effort, a key variable in much of the off-farm labor and contracting literatures. Differentiating between adult and child labor on the farm is another important dimension of labor input quality. National surveys of child labor are often conducted as standalone surveys, rather than integrated into agricultural surveys. As much child labor is crop-specific, detailed child labor data is often difficult to collect in national surveys. Input quality is also not limited to labor, as chemical inputs often face questions of quality due to inappropriate mixing, as in the case of pesticide use/exposure and fertilizer which may be adulterated (Michaelson et al., 2021, Norton et al., 2020). Asset ownership and use also vary considerably by respondent type (Doss and Quisumbing, 2019). We note advances in measuring labor and assets in section five.

4.2. The agricultural household model

A second class of models frequently used in agricultural economics link agricultural production decisions with household welfare. We sketch an agricultural household model to motivate measurement error and coverage biases when welfare analysis of production decisions is an empirical objective. In cases where separability is assumed, the model reduces to a profit maximization and utility maximization problem given production choices. The agricultural household model is a useful example of trade-offs in measurement error and coverage because we implicitly cover a range of producer models within the agricultural household model.

Household decisions are constrained by an agricultural production function, time endowment, and an intertemporal budget constraint (see Bardhan and Udry, 1999; LaFave et al., 2013; Singh et al., 1986). The household's problem is to choose own-produced agricultural goods (x_{at}), purchased market goods (x_{mt}), agricultural inputs (V_t), and leisure (l_t) to maximize the discounted stream of expected utility, given observed (μ_t), and unobserved household characteristics (ε_t).

In a non-separable formulation of the agricultural household model, production factors such as input prices also influence the household's consumption choices. Coverage biases may exist in the collection of input price data if household surveys do not measure market prices faced by farmers.

Imputed input price data for fertilizer, seed, or pesticides/herbicides ignore substantial price variation within input class correlated with product quality and efficacy.

Equation 1 provides the reduced form purchased market goods demand, which can be derived from the first order condition:

$$x_{mt} = x_{ct}(p_{mvt}, p_{avt}, w_{vt}, r_{t+1}, \pi_t(p_{vvt}, p_{avt}, p_{vt}, p_{At}; \theta_t), p_{vvt}, p_{avt}, p_{vt}, p_{At}, y_{vt}, \lambda_{vt}; \mu_t, \varepsilon_t) \quad (1)$$

where good m consumption depends on market (p_{mvt}) and own produced agricultural good prices (p_{avt}), the price of variable inputs (p_{vt}) such as agricultural labor, fertilizer, pesticides, or herbicides, interest rates (r_{t+1}), farm profits (π_t) conditional on climate variability (θ_t), exogenous income (y_{vt}), and future prices via the marginal utility of wealth (λ_{vt}). Consumption also depends on household characteristics, both observed (size and composition) and unobservable (food preferences). Input prices affect household consumption when markets are incomplete, and we cannot assume that income alone affects household consumption demand. Therefore, the consumption demand equation includes not only variables that affect household income, but also those that affect production decisions.

While we have discussed above the challenges in measuring agricultural variables in the demand equation, we now highlight coverage biases and measurement error in the estimation of equation 1. First, coverage biases could have significant effects on consumption demand when consumption is measured substantially after agricultural variables are realized and/or uses a different reference period. For example, annual household surveys that record production data from the last agricultural season may be lagged by months relative to household food consumption data, which is often recorded for the last seven-day reference period. Second, measurement error in food consumption aggregates can be substantial, due to the conversion of non-standard units and the subsequent imputed food prices (Oseni et al., 2017). Deaton and Zaidi (2002) provide a detailed description of consumption aggregate choices. As documented by Beegle et al. (2012), survey design choices related to different recall periods and survey modes (such as diaries versus in-person recall) have substantial effects on measured household consumption and consequently on imputed food prices and welfare. Third, an important specification issue in the demand equation is the inclusion of prices of consumption goods, own-produced goods, and inputs. As agricultural

surveys are often collected at a single period of time, capturing the relevant prices to correctly specify equation 1 could result in substantial measurement error given seasonal price fluctuations in both inputs and outputs.

In the next section, we discuss advances in measuring agricultural variables that are paramount to the producer and agricultural household models described above, but also to a wide set of models in agricultural economics that are beyond the scope of this paper. We note that the two models chosen in this section are examples, but issues of identification, measurement error, and coverage are not limited to producer and agricultural household models. Advances in data infrastructure improve internal and external validity by expanding the possibilities for improved identification and coverage, providing data sources for testing a wide range of potential empirical models.

5. Advances in Data Collection

The combined availability of new data sources, affordable computing power and data storage options, and digital technologies allowing for innovative modes of data collection (such as mobile and smart phones, tablets, and sensors of all kinds) have created a new data landscape with novel opportunities for more accurate, affordable, and timely data collection (Hill et al., 2019). In some cases, new data collection modes or innovations may help correct for existing biases – for example, measuring land area using GPS alongside farmers’ self-reported information – while in others, they may introduce new biases via under-coverage or response biases – for example, phone surveys or citizen-generated data (Amaya et al., 2019; Hill et al., 2019). Integration of new data collection modes with household surveys requires assessing trade-offs between cost, measurement error, coverage bias, and knowledge generated from testing new empirical models. A recent surge in the number of survey experiments, including on issues related to agriculture and food, has greatly contributed to making rigorously evaluated progress in survey design in areas of interest to agricultural economists (De Weerd et al., 2020).

In sections 2, 3 and 4 of this paper, we covered measurement error, coverage bias, and provide some examples of the expanded frontier of empirical models. In this section, we discuss advances in data collection with a focus on their impact on reducing measurement error and/or increasing

coverage.⁴ With this context in mind, we organize this section around (1) advances in specific thematic areas of relevance to agricultural economists, and (2) modes or data structures that provide new solutions and address challenges to reducing error and increasing coverage. For both topics, we highlight how these advances speak to the issues highlighted in the previous sections, including Total Survey Error, bias, measurement error, coverage, inter alia. Many recent advances in data collection have resulted from addressing constraints to data collection in low- and middle-income countries, but we also highlight experiences from high-income settings to emphasize how these issues are in fact globally relevant.

5.1. Advances in selected thematic areas

Land area measurement

Recent evidence from studies in Africa (Abay et al., 2019a; Carletto et al. 2017b, 2016; Desiere and Jolliffe, 2018) and Asia (Dillon and Rao, 2021) that included both GPS and self-reported measures of **land area**, all following the type of survey experiment set-up advocated in De Weerd et al. (2020), found remarkably consistent presence and patterns of non-classical measurement error in farmers' self-reporting. The superiority of GPS with respect to self-reported measures has been confirmed by studies that also included the more expensive compass-and-rope method, such as Carletto et al. (2016) and Dillon et al. (2019). The integration of hand-held GPS devices in survey work for land area measurement has since become commonplace to overcome this prototypical example of respondent effect. While GPS measurements are not entirely free of error (Dillon et al., 2019; Cohen, 2019), the associated measurement error is larger in relative terms for very small plots but is not correlated with land size (Carletto et al., 2017b, 2016).

Innovation is now proceeding in the direction of integrating GPS measurement in CAPI applications through the testing of features allowing plot delineation on preloaded satellite imagery (Masuda et al., 2020) or on printed high resolution imagery (Dillon and Rao, 2021), or through the use of the GPS receivers integrated into interviewer tablets for *in situ* land area measurement. In geographies where accurate knowledge of land area by respondents is commonplace, similar

⁴ A more comprehensive, and prescriptive, treatment of agricultural survey design choices is provided in Dillon et al. (2021a), which builds on the guidance in Glewwe and Reardon (2000).

technological developments are being pursued to enable the efficient delivery and evaluation of programs tied to land area and land use, such as the Land Parcel Identification System (LPIS) in the European Union (Pluto-Kossakowska et al., 2008; Devos 2011; Tarko et al., 2015), or in the area survey implemented by the National Agricultural Statistics Service (NASS) in the United States, which is also successfully experimenting with the use of a mobile plot delineation application (Abreu et al., 2017).

In the coming years, these developments can be expected to be brought to scale to address some of the drawbacks of measuring land area with GPS units, such as the cost of plot visits and the inability to measure all plots, particularly those that are more distant or particularly large. While *in situ* GPS measurement certainly reduces bias, some of these concerns about item non-response can be mitigated through imputation methods, which have been shown to effectively predict GPS plot measures for all plots by using farmers' self-reporting alongside other plot characteristics (Kilic et al., 2017), or by further technological development if plot delineation on high resolution imagery can reduce the drudgery of the field visit that typically plagues GPS measurement.

Agricultural output and yields

Recent empirical work has reviewed the quality of agricultural **output data**, related to both the level of data collection as well as biases in farmers' self-reporting of agricultural output. Abay et al. (2019a), Desiere and Jolliffe (2018), Gourlay et al. (2019), and Lobell et al. (2020) all point to the presence of non-classical measurement error in farmer's self-reporting of crop output, with farmers substantially over-reporting production on small plots and under-reporting production on larger plots. Currently, these biases can be corrected through the use of crop cuts on sub-samples, and looking ahead, through Earth Observation data calibrated with ground-truthing from field observations. Two levels of integration will be key to moving the agenda forward: integration between subjective (recall) and objective (crop cuts) data, and between ground and satellite data. Where available, administrative data can also be combined with survey data (as well as with satellite imagery and climate data) to produce disaggregated model-based yield estimates (see for instance Erciulescu et al., 2019 for county-level yield estimates in the United States).

Meanwhile, challenges persist in the measurement of **yields** in fields using mixed or inter-cropping planting techniques (Dillon et al., 2020; Wineman et al.; 2018). Estimating land area apportioned to a specific crop as well as its production is particularly difficult. Most household surveys acknowledge the complications of production and input estimates on inter-cropped plots by identifying these plots and apportioning the area planted, to divide reports of plot-level inputs by production reported by crop. However, proportional input attribution implies crop input demands including fertilizer, weeding, and harvest time are similar by crop, which may not always be an accurate assumption.

The Global Strategy to Improve Agricultural and Rural Statistics provides methodological guidance on implementing the above methods to measure the area under a given crop in inter-cropped systems (GSARS, 2018). Unfortunately, guidance on best practices supported by evidence from methodological survey experiments is not currently available. Remote sensing or crop cut production estimates are possible alternatives, but these measures are also challenging to implement. For instance, crop cutting, in addition to its high costs due to the need for closer supervision and multiple visits over the growing period, can only be done in a very restricted time window which may be difficult to plan correctly in a large survey operation. It also carries implementation difficulties that are associated with specific error generation mechanisms (Kosmowski et al., 2021). Furthermore, Wahab (2020) find a substantial discrepancy between crop-cuts and self-reported output measures, which he ascribes in part to the variability in crop performance within plots, leading to plot area loss in the course of the season.

Yield prediction models based on remote sensing data clearly face bigger challenges the smaller the plots and the more complex the cropping patterns, particularly related to the degree of intercropping or the presence of canopy cover. Lobell et al. (2019) report lower accuracy of remotely sensed production estimates compared to crop cut production estimates for maize inter-cropped plots in Uganda. However, they also clearly show the benefit of properly calibrating the spatial model through accurate ground-truthing based on high-quality crop cutting, even if only on a small sub-sample of plots. Řezník et al. (2020) compare yield predictions from satellite data with measured yield data on spring barley, winter wheat, corn, and oilseed rape in the Czech Republic,

finding the yield predictions to be credible, with only two out of nine measures reporting differences between measured and predicted yields larger than 5 percent.

Agricultural labor

SDG 2.3, which defines productivity in terms of output per unit of labor, has increased attention to the measurement of **labor** productivity. At the same time, results from survey methods research have unearthed the staggering magnitude of recall bias (a respondent effect) in measures of agricultural labor, with one influential study showing hours worked per person-plot being 3.0 and 3.7 times higher in recall surveys compared to benchmark estimates based on weekly visits (Arthi et al., 2018).

Agricultural labor data have been typically sourced through labor force surveys or national censuses (with information generally limited to the primary occupation) and used primarily in aggregate-level productivity analysis and macro-level comparisons of national agricultural GDP with labor shares. The availability of higher quality labor data in the last decade has raised questions about the validity of evidence that shows a six-fold labor productivity gap between agriculture and non-agricultural sectors of the economy (Gollin et al., 2014). Studies that use more carefully collected labor data from household surveys have shown that the measured labor productivity gap is substantially reduced when data allow for measuring production per hours worked, as opposed to just per person per year (McCullough, 2017), and for individual fixed effects (Hamory et al., 2021). In the US, where data on agricultural labor are collected via a dedicated survey, farm labor hours have historically been difficult for respondents to report, as a low percentage of operators based their responses on formal records (National Research Council, 2008; Ott, 1999). Difficulties in this case refer also to capturing the complexity of the pay structure, recording information on different tasks, since many agricultural workers perform multiple tasks on the farm (Ridolfo and Ott, 2021), and collecting data on contract workers (Ridolfo and Ott, 2020).

Advances in the measurement of labor inputs in recent years have been based on both technology-enhanced and low-tech innovations, including by leveraging mode of data collection to ease the

cognitive response burden. Notable technology-enhanced innovations include the use of mobile phones for high-frequency interviews (Arthi et al., 2018; Dillon, 2012), and the use of wearable accelerometers for the measurement of physical effort (Akogun et al., 2020). Arthi et al. (2018) find that phone surveys can be a more accurate alternative to face-to-face interviews for measuring labor inputs, and this finding remains consistent when the research question calls for collecting high-frequency data or repeated measures. In such cases, the cost of additional phone interviews is a fraction of the cost that would be implied by additional face-to-face visits (Table 1).

Table 1. Per-Household Interviewing Cost Increases

# Interviews	Weekly Visit	Weekly Phone
1	14%	6%
10	139%	54%
20	277%	108%
25	346%	135%
30	416%	162%

Note: The costs are the cost increases in US Dollars, per household, relative to the cost of an LSMS-type (baseline) survey.

Source: Arthi et al., 2017.

Akogun et al. (2020) measure the physical activity of sugarcane cutters using accelerometers, which is a direct measure of effort in their piece rate wage setting. They find a high correlation between administrative data on output per worker recorded by the firm and the worker's physical activity, as well as large changes in the intensity of such activity in response to malaria testing and treatment. Integrating objective physical activity measures into a sub-sample of observations in national surveys may be used to calibrate biases in reported time as well as to predict effort-based measures of agricultural labor productivity.

Aside from the mode of data collection, substantial recent advances in methodologies relate to the key set of survey design choices in agricultural labor measurement. Bardasi et al. (2011) investigate how survey design elements such as screening questions and proxy response results in biased estimates of labor force participation, hours worked, and income by gender and sector of employment. Female labor participation statistics are not affected by the use of proxy respondents

in their survey experiment from Tanzania, but male employment rates are, due to the under-reporting of agricultural activity by proxy respondents. Using data from Malawi, Kilic et al. (2020) find that employment is further under-reported when recall periods increase and when women are the subject of proxy reporting. Recent advances in data collection software and the ubiquitous use of CAPI can also make it easier to avoid another source of coverage-related bias unearthed by Ambler et al. (2020). They show that the fact that household members are not listed randomly in the labor module, coupled with respondent fatigue, leads to age and gender related biases in employment measures. Software that allows for randomizing the ordering of household members when collecting data on the labor module can mitigate this source of systematic bias, as can avoiding the use of proxy respondents. Avoidance of proxy respondents to minimize measurement error, however, can potentially lead to greater errors of coverage.

The effects of different recall periods for measuring agricultural labor are investigated by Arthi et al. (2018), who use Tanzania data to compare weekly agricultural labor reporting with end-of-season reporting. The latter is associated with a fourfold increase in the hours reported by individuals at the plot level, in comparison to reports obtained via weekly visits, their preferred benchmark. However, they note that aggregation to household-level reporting causes the differences in reported hours between the weekly and end-of-season recall periods to disappear. In interpreting these findings, the authors note how different recall biases are associated with memory decay (which shorter recall would help address), but also by the mental burden of reporting that varies by the level of aggregation. In their study, aggregating plot-person hours to the household level happens to compensate for competing biases arising from over-reporting at the intensive margin and under-reporting at the extensive margin. However, this is not a result that can be extrapolated to other settings. Understanding the level of disaggregation at which individuals provide the most accurate reports on their agricultural labor inputs should be an area of focus for future research. Research by Gaddis et al. (2020) in Ghana find much less dramatic differences in the magnitude of the recall bias compared to Arthi et al. (2018), but also discover that an important source of bias is the omission of plots and farm workers at the listing stage, which can be mitigated by explicit attention to this specific aspect of survey design.

In the United States, a substantial amount of randomized testing (Reist et al., 2019) and cognitive interview piloting (Ridolfo and Ott, 2020; Ridolfo and Ott, 2021) is routinely devoted to testing innovations aimed at easing response burden and addressing complex questions about workers' remuneration and tasks. The findings suggest that the optimal design of instruments to collect labor data will likely require a fair amount of adaptation based on the context and the intended use of the data. For instance, while respondents in the United States appear comfortable separating base and overtime hours, they had difficulties distinguishing base pay from bonuses – the concept being hardly applicable to respondents paying piece rate (Ridolfo and Ott, 2021). For low-income settings, Sagesaka et al. (2020) have systematized recent findings from survey research into practical data collection guidance for survey practitioners.

Non-labor inputs

One empirical regularity that has recently come to the fore is that measurement error in land area is strongly correlated with farmers' self-reporting of their application levels of **agricultural inputs** (Abay et al., 2019b; Bevis and Barrett, 2019; Burke et al., 2019). These patterns in the data naturally raise questions on the mechanisms that drive the relationship between non-classical measurement error (NCME) for land area and self-reported input application rates. One such mechanism could be that farmers have a mental heuristic for input application rates and thus self-report, for example, seed or fertilizer quantities based on the amount of land they believe they cultivated, along the lines of the optimal error prediction model of measurement error. Such a heuristic is easy to imagine in the case of fertilizer or seed, for which extension agents and agricultural input dealers commonly offer recommendations in the form of application rates per unit of land cultivated. If this is indeed the mechanism behind the observed correlation between area NCME and agricultural input levels or application rates, it could imply either of two possibilities. On the one hand, NCME in land area might propagate into NCME in agricultural input data – that is, the measurement error in inputs would merely reflect the error in land area, permitting statistical correction using observed area measurement error. On the other hand, land area NCME could actually affect agricultural input use by farmers, if farmers' decisions on input intensity are based on misperceived land area (Abay et al., 2019b). Eliciting input use information after the collection of objective land area measures to better understand how the mental heuristic

of optimal application rates may be influencing farmers' self-reporting is a key methodological research area for improving data collection on input use.

Aside from application rates, measuring the quality of inputs is an important and often unobserved characteristic of input investments. The fact that input quality is often not directly observable poses a problem not only for the analysis of agricultural productivity, but also for farmers in making decisions on input use. Perceived quality may influence input demand and use more than actual attributes of quality. Such questions have been difficult to explore until recently, as economists have begun complementing traditional data collection from farmer respondents with laboratory analysis. The latter is also not free from error, however. An early study by Bold et al. (2017) finding widespread problems with nutrient quality in Uganda has since been contradicted by a series of large-scale sample surveys finding limited evidence of widespread quality issues in synthetic urea in East Africa. There is also evidence that perceptions of quality are influenced by other factors that in turn influence productivity, such as rainfall patterns (Hoel et al., 2021; Michaelson et al., 2020; Ashour et al., 2019a; Sanabria et al., 2018). Collecting better data on both perceived and actual fertilizer quality is essential to explain farmers' behavior with respect to their adoption, and the extent to which possible remedial action for low levels of fertilizer use may come from certification or the use of other policy levers (Hoel et al., 2021).

For herbicides, Ashour et al. (2019b) find that there are widespread quality issues with the herbicides available in local markets in Uganda, but that farmers' perceptions of poor herbicide quality are overstated, and poorly correlated with actual measures of product quality from laboratory testing. Prices correlate with measured quality, but very weakly. In a technical report using the same data set, Ashour et al (2019a) report poor correlation between tests in two different labs and ascribe the difference to flawed procedures in one of the facilities, a reminder to researchers that 'objective' measures conducted with the aid of technology are, as with any measurement operation, not immune from error.

In countries that have administrative data systems around the use of agricultural inputs such as pesticides, these offer the potential to be combined with survey data to both improve the accuracy of the data compared to respondents' recall, while also reducing the burden on survey participants.

This is for instance the case in the United States, where at least some states (Arizona, California) are using data from mandatory pesticide use reporting systems instead of asking farmers (NRC, 2008). However, these methods may be more difficult to implement when the objective is to collect crop or field level data: in such cases, the US National Agricultural Statistical Service (NASS), collects data from respondents on one randomly selected field for selected crops of interest (NASS, 2021). A similar type of use of multiple data sources may also be more difficult to implement in poorer countries, where administrative data systems suffer from low quality and credibility.

These studies are examples of ways in which administrative or market-level data collection can be combined with household-level survey data to provide evidence on the use and quality of inputs available to farmers. In terms of our conceptual framework, this implies efforts towards improving the accuracy of input quality (via objective testing) and quantity (via the use of administrative records) estimates as well as the coverage (e.g. via market-level sampling for quality testing which can be linked to farm level behavioral variables), but also to the collection of additional (omitted) variables related to farmers' perception of quality, as these may be only tenuously linked to actual quality attributes.

Soil quality and soil health

Stevens (2018) writes that **soil health** “is a straightforward concept in the abstract, but difficult to define in practice”. Not only do soils have many attributes that require multiple, complex measures, but these attributes are also interdependent, and the attributes (or their combinations) of significance can vary depending upon the application for which an assessment of soil health is needed.

In Europe, the ‘Land Use/Cover Area Frame Statistical Survey Soil’ (LUCAS Soil) is a regular topsoil survey that is implemented every three years on approximately 20,000 soil samples collected across the European Union (Orgiazzi et al., 2018). The United States Department of Agriculture’s Natural Resource Conservation Service (NRCS) maintains a century-old soil survey of the United States (NRCS, 2021). While both these data sets have relatively good national coverage and are spatially explicit, their use in conjunction with the main farm surveys in the

European Union and United States for economic and policy analysis remains limited, partly due to the difficulty of overcoming confidentiality concerns in data dissemination preventing record linkage across data sets (NRC, 2008).

In low-income settings, where large-scale soil surveys are not usually available, recent research has cast serious doubts on the reliability of farmers' self-reporting on soil quality and soil health, with findings for Ethiopia (Carletto et al., 2017a; Kosmowski et al., 2020a), Kenya, and Tanzania (Berazneva et al., 2018) consistently finding poor or no correlation between farmers' assessments of soil quality and objective measures based on lab analyses or portable spectrometers. Unlike land area measurement, there are no clear systematic biases emerging in the case of soil quality attributes; the concern is mainly with the lack of explanatory power of the traditional measures relying on farmers' assessments. While some predictive power has been reported for soil type (Berazneva et al., 2018) and soil color and texture (Gourlay, 2017), the reported correlations are very weak.

Efforts to pilot the use of portable spectrometers for *in situ* objective measurement of key soil health features such as organic carbon, PH, nitrogen, potassium, and clay percentage have been shown to perform well when compared to Conventional Soil Analysis (Carletto et al., 2017a; Kosmowski et al., 2020; Vasques et al., 2020). While portable spectrometers are not nearly as widely available as GPS units, their cost and weight are expected to decline rapidly as technology advances, making the prospects for their use at scale ever more attractive, particularly when soil attributes are important for the research question at hand. In lieu of field-ready soil sensors, some survey efforts have moved towards smartphone-based soil assessments such as LandPKS (Herrick et al., 2013), but these have largely been on pilot-level or small-sample surveys (see for example Nord and Snapp, 2020).

The other related avenue through which advances in soil health data can be expected to rapidly materialize is the integration of remote sensing data with georeferenced survey data. The correlation between available modeled georeferenced data such as AfSIS (see Hengl et al., 2015 for details) has been shown to be encouraging but far from perfect, particularly when there are high variations in soil quality within a given geography (Gourlay et al., 2017). As more objectively

measured ground data on soil health is collected and used to train models based on Earth Observation data, however, the quality of the modeled data will increase (Kosmowski et al., 2020a).

Agricultural machinery and farm implements

Agricultural capital in the form of **machinery and farm implements** can increase the production capacity of smallholder farmers. Understanding the mechanization of agriculture is critical to understanding changes in farm size and profitability over time. While it is generally regarded as easy for farmers to recall agricultural capital within the household, the plot-level attribution and control of such capital are measurement challenges. Plot-level attribution of machinery use is often avoided, as it may be assumed by the survey designer that agricultural capital is shared equally in the household.

A large literature on women's empowerment in agriculture has focused on accurately measuring women and men's ownership of assets relative to their use rights (Alkire et al., 2013). Doss and Kieran (2014) provide a comprehensive review and guidelines for collecting gender-disaggregated asset data which apply generally to agricultural capital modules. Kilic and Moylan (2016) provide experimental evidence on the effects of variation in respondent selection protocols and questionnaire design compared to commonly used approaches for eliciting information on the individual ownership of and rights to assets. These studies all emphasize the importance of respondent selection and the method of collecting ownership and use rights. Lessons learned from this body of work have been consolidated in the recently published *Guidelines for Producing Statistics on Asset Ownership from a Gender Perspective* (United Nations, 2019). Data from the machinery and farm implements modules can be linked to plot-disaggregated production and other inputs modules to assess differences in the intra-household allocation of inputs (Udry, 1996).

Recall periods for agricultural machinery and implements usually focus on the availability of assets over the previous 12 months. Differences in input use by crop-plot-season are important to capture, but this may not be possible if the frequency of survey administration is annual rather than

seasonal. The age of machinery is usually collected with the intention of calculating depreciation, but much depreciation of machinery depends on their maintenance and frequency of use.

Crop variety identification

Possibly the most important technological choice farmers face is that of choosing which crop, and specifically which **crop variety**, to plant. A good proportion of the budget for agricultural research globally is directed at breeding crops and livestock with desirable traits. While the uptake and impact of improved varieties has traditionally been collected by eliciting information from either farmers or panels of experts, the shortcomings of such methods have become evident in the past decade; as a result, they are gradually being replaced or combined with more objective methods (Maredia et al., 2016; Stevenson et al., 2018; Wossen et al., 2019). The method that is currently being more widely adopted is DNA fingerprinting, which entails the collection of plant material that is subsequently sent for lab analysis. While logistically cumbersome, its implementation has been shown to be possible at reasonable scale, and protocols for its adoption are emerging (Poets et al., 2020).

Asking farmers to identify the crop variety they are planting has often been shown to be utterly inaccurate, even when augmented with photo aids or phenotypic trait-related questions aimed to improve the accuracy of the data. This holds true for different crops across different settings, including sweet potato (Kosmowski et al., 2019), wheat, maize, barley and sorghum in Ethiopia (Jaleta et al., 2020; Kosmowski et al., 2020b; Yirga et al., 2016), cassava in Ghana and beans in Zambia (Maredia et al., 2016), maize in Uganda (Kilic et al., 2017) and Tanzania (Wineman et al., 2020), and cassava in Vietnam (Le et al., 2019), Colombia (Floro et al., 2017) and Nigeria (Wossen et al., 2019). A few studies report more encouraging self-reported results, with farmers in Bangladesh being most able to discern modern from traditional varieties for both rice (Kletzschmar et al., 2018) and lentils (Yigezu et al., 2019). The latter study is also of interest in that the panel of experts was, on the contrary, found to overestimate adoption by 89 percent compared to DNA fingerprinting.

Even in the studies where farmers' self-reporting is close to the objective benchmark, DNA fingerprinting was found to have advantages for the analysis of determinants of adoption (Yigezu et al., 2019) as well as for detecting lack of authenticity in modern varieties present in seed markets and in the field (Kletzschmar et al., 2018). When technology adoption is an important component of research design, researchers should consider adopting DNA fingerprinting as a data collection method. The option of conducting such objective, yet more costly, measurement could be more routinely considered on a sub-sample or for priority crops of interest. When field visits for area measurement or crop cuts for output measurement are being performed, the research design can exploit significant economies of scale by performing additional tasks during the same visit to the plot. This does pose other constraints to data collection processes, as such field work needs to be performed within a specific time window (i.e., while crops are still in the field). Ethiopia has been able to incorporate DNA fingerprinting at scale in a national socioeconomic survey for three main crops: wheat, barley, and sorghum (Kosmowski et al., 2020b). Barriga and Fiala (2020) use DNA lab analysis to investigate seed quality along the seed supply chain, looking at genetic variation, physical purity, and performance, focusing for the latter on germination rate, moisture level, and vigor. This allows them to identify issues with the handling and storage of seeds, rather than counterfeiting or adulteration.

In addition, Kosmowski and Worku (2018) report promising results for the use of spectrometers for varietal identification on cultivars of barley, chickpea, and sorghum in Ethiopia, with an overall correct classification accuracy of respectively 89, 96, and 87 percent in their sample. Sinha et al. (2020) report similarly encouraging results from a study on banana varieties in Uganda by extrapolating the ground-based hyperspectral measures to high-resolution satellite imagery, therefore creating the potential of mapping the distribution of banana varieties at a higher spatial resolution. This is an exciting area of innovation which is currently at the experimental stage but is likely to become mainstream over the next few years, provided validation efforts continue and implementation protocols are devised.

Measurement of farm level food losses

While research on **food losses** has increased in recent years, the available data is extremely heterogeneous with respect to the measurement approaches used, the stages of the value chain investigated, and the conceptual framework adopted. Bellemare et al. (2017) propose a different conceptualization of food waste from that used by others in this domain, whose estimates of food losses would be largely overestimated according to their definition (Table 2).

Table 2. A Comparison of Quantity and Cost Estimates of Food Waste Across Definitions

Food Waste Definition	Quantity Estimate	Cost Estimate	Cost Estimate Using Our Cost Proposal
Our Definition	42.8	27.6	—
EPA	20.0	—	17.8
FAO	57.6	39.6	37.1
ERS	57.6	57.6	37.1
FUSIONS	69.1	47.8	44.6

Source : Bellemare et al., 2017.

In the existing literature, storage is the stage of the value chain where most food losses are concentrated (FAO, 2019).⁵ Xue et al. (2017) attributes differences in food losses to different storage conditions, and research from Bachewe et al. (2018) and Minten et al. (2015) also point to the importance of storage losses. Despite the interest and prominence that the debate on food losses has acquired, data of sufficient quality and robustness on storage losses is lacking, hindering the design and implementation of interventions to reduce them systematically and at scale.

Comparisons between objective and self-reported measurements of food losses routinely find systematic differences between the two. While objective measures are more accurate, they are also more costly, time-consuming (selecting, sorting, and weighing samples of grains), and logistically challenging. Model-generated methods of estimation are therefore being researched, as they offer a possibility to deliver measurements in a more cost-effective manner (FLW Protocol Steering Committee, 2016). Model-based estimates could be used in conjunction with rather than as a replacement for survey data, for instance, by estimating losses between survey rounds. These

⁵ The discussion on food losses below draws on text provided by Marco Tiberti and FAO, based on unpublished material.

estimates can determine storage outcomes by taking into account the effect of variables related directly to storage conditions (e.g., the type of storage facility, the application of pest protection products, or the moisture content at which the grain is stored) as well as contextual variables (e.g., weather conditions, crop variety, or farmer skills), and the interaction between the two. The African Postharvest Losses Information System (APHLIS) is one example of the production of losses estimates based on the modeling of agronomic and bio-physical relationships of factors including the presence of rain at harvest time, as well as agricultural storage and marketing practices.

Livestock production and management

The bias of agricultural economists for crops over **livestock** is reflected in the relatively limited efforts seen to date on developing better data collection methods for livestock (Barrett et al., 2008; Kristjanson et al., 2014; Little et al., 2008; McCarthy et al., 2004; Pica-Ciamarra et al., 2014). Most methodological work has been directed at pastoral or agro-pastoral systems, which is to be expected, given both the specific challenges these systems pose to data collection and the importance of livestock for people living in regions where pastoralism is prevalent. Recent work in this area has focused on herd mobility, to address the challenges that it poses for enumerating nomadic or semi-nomadic populations, as well as to study mobility patterns linked to the state and management of natural resources (e.g., grazing, water) upon which livestock and their herders depend.

For example, Himelein et al. (2014) conducted a pilot in the Afar region of Ethiopia to explore the use of random geographic cluster sampling as an alternative to conventional sampling methods. The approach is based on the random selection of points around which circles are drawn and all eligible respondents found inside those circles interviewed. The approach aims to reduce the under-coverage of mobile populations expected when samples are drawn based on lists of dwellings within a primary sampling unit, as is typically the case for household surveys. Otherwise, methods have not evolved significantly from the surveys at enumeration points (i.e., water, dipping or vaccination points, stock routes) and aerial surveys recommended by ILCA in the 1970s and 1980s (FAO, 1992; GSARS, 2016; ILCA, 1990), except for the fact that these aerial

surveys can now also be implemented using higher-resolution imagery captured by drones (Chamoso et al., 2014) or satellites. However, these methods are still in the experimental stage and have not to our knowledge been applied at scale.

Advances in spatial data, both from satellites and on the ground, is creating opportunities for the collection of data on the interaction between livestock, mobility, and natural resources. On the ground, GPS trackers placed on cattle have been used to characterize the mobility of herds and their use of rangeland resources (Bailey et al., 2018; Liao et al., 2018, 2017; Swain et al., 2011; Turner et al., 2000), although few of these applications have appeared in economics journals. From space, satellite imagery is being used to characterize the state of rangeland resources (Reinermann et al., 2020) and we expect that the potential for applications in agricultural and natural resource economics will expand dramatically as a result.

On improved measures of livestock productivity, recent studies led by economists are limited. Specialized livestock surveys often select a random animal in the herd and ask questions about that animal. In household surveys, this is not generally done, as the herd may not be present, and a visit would add to the interview time. Livestock experts also tend to measure productivity using the reproductive capacity of the herd, and thus their focus is on demographic parameters (Lesnoff et al., 2014). For milk off-take, a methodological study conducted in Niger comparing different types of recall to an objective measure provides some confidence in the accuracy of recall measures (Zezza et al., 2016a). Other technologies, such as 3D and thermal cameras, are being used to assess livestock weight and health (Song et al., 2018; Stajanko et al., 2008), but mostly by animal scientists rather than economists or statisticians. Nonetheless, there is a clear potential for economic applications to emerge, as the value of livestock is primarily determined by parameters linked to weight and health, which are notoriously difficult to elicit from survey respondents. Guidance for data collection on livestock in low-income countries has been systematized in recent years in GSARS (2016) and Zezza et al. (2016b). Model-based estimates of livestock populations have been developed by researchers at the FAO (Robinson et al., 2014) and are continuously being updated as new spatial data sets become available and modeling techniques evolve (Nicolas et al., 2016; Da Re et al., 2020).

Land tenure

Holden et al. (2016) document that few low- and middle-income countries have nationally representative data that can be used to understand how **land tenure** policies or tenure reforms may affect land market activity, land productivity, technology adoption, or changes in the distribution of farm size. Measurement challenges in this area are primarily related to the complexity of the concept of tenure and the different set of rights that define it (FAO, 2002; United Nations, 2019), as well as to the fact that different individuals may have different perceptions of tenure, particularly in the case of joint ownership (Ambler et al., 2020; Kilic and Moylan, 2016). In high-income countries, increasing challenges for data collection arise for more complex forms of ownership, linked to the rise of corporate land ownership and of complex company arrangements for corporate farms (National Academies of Sciences, Engineering, and Medicine, 2019; MacDonald, 2016).

With respect to adequately capturing the different dimensions of tenure, the consensus has converged towards the need for survey data to cover a bundle of ownership rights, including documented ownership, reported ownership, and the rights to sell and bequeath (United Nations, 2019), and survey instruments have been developed to implement this guidance (FAO et al., 2019). In the United States, where corporate farms account for an increasingly important share of agricultural value added, the Agricultural Census form includes a specific set of questions on the type of farm organization (whether a Limited Liability Company) and their legal tax status (family, partnership, incorporated, or other). The census information is then integrated with a separate Tenure, Ownership, and Transition of Agricultural Land (TOTAL) Survey, which focuses specifically on all land rented out for agricultural purposes, whether by farmers and ranchers (operator landlords) or by non-operator landlords. Given the complexity of some operations, the surveys face challenges in the definition of the landlord, the identification of the owners particularly when incorporated, and in assessing the location and size (combined acreage) of landowners (Hamer, 2016).

For household farms, whenever individual-level data are of interest, such as when the research objective is to study gender gaps in productivity, wealth, or vulnerability, land ownership should be reported by self rather than proxy respondents, owing to well-documented and large

discrepancies between proxy and self-responses. While research on the implication of different possible approaches is still needed, the primary issue is the method of respondent selection, where researchers increasingly favor interviewing multiple individuals per household. Approaches may vary, and will also depend on the objective of the analysis, but they can be reduced to essentially three options: (1) interview all household members, (2) focus on the members of the principal couple if one is present, or (3) select a random age-eligible household member and his/her partner if applicable (Doss et al., 2019). When multiple household members are interviewed, they should be interviewed separately and whenever possible concurrently or consecutively, so as to avoid the possibility of contamination in their responses (United Nations, 2019).

Climate: weather events, perceptions of and adaptation to climate change

Climate data have experienced a revolution in recent decades which continues to the present day. While climate and weather have always been central to explanations of agricultural productivity, attention has increased with the emergence of debates on climate change, climate-smart agriculture (Lipper et al., 2018), and index-based insurance (Benami et al., 2021; Carter et al., 2017; Jensen and Barrett, 2017; Rosenzweig and Udry, 2014). Dell et al. (2014) and Auffhammer et al. (2013) provide excellent reviews of the types of available climate data as well as their accompanying measurement bias and coverage concerns, which economists should consider when relying on climate data for making inferences. In terms of the production and availability of climate data, there has been a surge in data from remote sensing and *in situ* sensors (which are discussed later in the paper), as well as concerns in Africa and small island states regarding the decline in the availability of traditional meteorological stations (Dinku, 2019; Dobardzic et al., 2019).

Weather data are commonly classified into four categories: ground station data, gridded data, satellite data, and reanalysis data. Data from ground stations offer direct observation of key weather variables, but their coverage is neither universal nor constant over time, with weather stations being relatively sparse in many low- and middle-income countries. Additionally, their coverage and trends are often related to the distribution of weather variables, posing estimation problems similar to those of selective attrition. Gridded data provide complete coverage at different resolutions by interpolating weather station data and assigning a value for weather

variables for each cell on the grid. They present the desirable advantages of balanced panels, but analysts should be aware that results will differ for different products, particularly for outcomes that have greater spatial variation such as precipitation. The presence of missing values in the underlying station data and the spatial correlation introduced by extrapolation algorithms all create potential biases in the estimated coefficients and standard errors when gridded products are used as independent variables in econometric analyses (Dell et al., 2014).

Satellite data use readings from satellite-borne sensors but do not directly measure weather events. Their time series are shorter than those for station and gridded data (starting in the 1990s and increasing since the 2000s), and their quality may not be uniform, due to changes in satellites and sensor features. Reanalysis data combine information from other weather data sources and elaborate them with a climate model to estimate (not simply interpolate) weather variables across a grid. Analysts should consider whether such modeled data are preferable to interpolated gridded data, given the objective of the analysis, and should be aware that the correlation across models is often weak, particularly for rainfall data. Dell et al. (2014) and Auffhammer et al. (2013) provide a more detailed discussion, while Michler et al. (2020) and Parkes et al. (2019) provide examples of empirical applications testing the behavior of different gridded products as explanatory variables in agricultural productivity analyses for India and Sub-Saharan Africa, respectively.

Analysts must also identify the most appropriate set of climatic variables to use when specifying explanatory models for outcomes heavily dependent on climatic inputs. Advances have come from the increased cross-fertilization between crop science and statistical models, which has expanded the range of climate variables used in empirical analysis beyond standard rainfall and temperature measures. Newly included climate variables include growing degree days (GDD) and extreme heat degree days (EHDD) as well as measures to better account for humidity and evapotranspiration such as vapor pressure deficit (VPD), wind speed, and sunshine duration (Roberts et al., 2013; Zhang et al., 2017). Challenges remain for statistical models in accounting for the effects of carbon dioxide (CO₂) that accompany warming or the concentration of ozone (O₃) that may be associated with the burning of fossil fuels (Lobell and Asseng, 2017).

In parallel, there have also been increased efforts to capture both subjective perceptions of climate change, as well as practices reflecting the adoption of adaptation practices by farmers. While several researchers have engaged in collecting data with this objective in mind (e.g., Di Falco, 2011) there have been few attempts (McCarthy, 2011) to systematize data collection instruments in this domain; as such, this remains an area in need of further development. However, recent studies comparing self-reported data on weather events to recorded, observed weather data find a very weak correlation between the two. More importantly, they find that self-reported weather data are influenced by variables of interest such as the involvement in off-farm activities (Nguyen & Nguyen, 2020; Waldman et al., 2019). Self-reported data hold more promise for investigating perceptions and adaptation actions by farmers, whereas indicators referring to realized weather events should be based on objective data whenever possible. Researchers of smallholder, rain-fed production systems face particular challenges in achieving the granular resolution required for conducting plot-level analysis of the determinants of productivity, yield variability, and other key outcomes.

5.2. Advances in data collection modes and data structures

Earth Observation

The ever-increasing number of satellites orbiting the Earth has exponentially increased the availability of satellite-borne sensors supplying a variety of data at high temporal and spatial resolution. A classification of the satellite sensors categories, with their main features, as per the classification of the European Space Agency is provided in Table 3.

Table 3: Classification of satellite sensor categories, based on the European Space Agency (ESA) nomenclature

Mission group	Resolution	Examples	Utility in context of agricultural statistics
SAR	High, 4–30 m	Sentinel-1 Radarsat-2 ALOS-PALSAR 2 RISAT	6- to 24-day revisit, C-band (S1, R2) and L-band (ALOS). All weather, day and night imaging, including interferometry. Free (S1) and fee-based (R2, ALOS, RISAT). Contribute to crop delineation and identification (< 20 m resolution).
	Very high, 1–4 m	TerraSAR-X CosmoSkyMed Radarsat-2 (fine mode)	Multiple sensors, variable revisit (up to daily frequency). High-resolution but small swath. X-band (TSX, CSM) is less suitable for crop identification. All can acquire both high- and lower-resolution imagery, exclusively. Fee-based, expensive.
Multispectral	High, 4–30 m	Landsat, Sentinel-2 ASTER SPOT 6/7 RapidEye CBERS IRS LISS DMC	Free and open access to 10–30 m (Landsat, S2, ASTER), with large swath and including Short-Wave Infrared (SWIR) channels. 5- to 16-day revisit or better (e.g. RapidEye). SPOT combines 6 m multispectral with 1.5 m panchromatic. Suitable for crop delineation, crop type identification and area estimates (4–10 m resolution), regional crop occurrence and status (4–30 m).
	Very high, 0.3–4 m	WorldView-3 Pleiades PlanetLabs SkySAT DMC-III	Highest resolution (0.3 m WV-3), variable number of spectral bands, in visual and near-infrared (NIR) and SWIR range. Entry of new suppliers (e.g. PlanetLabs, SkySAT) is driving down (currently steep) prices. Suitable for parcel measurement (< 1 m resolution).
	Medium, 30–300 m	Proba-V Sentinel-3 MODIS VIIRS	1–3 days revisit over large swath (> 500 km). Including thermal channels (MODIS, VIIRS, S3). Composites for periodic trend analysis over large areas. Parcel detail typically lost. Useful for typification of large production areas and seasonal trends, for stratification. Free and open access except Proba-V (100 m resolution).
	Low, > 300 m	MODIS AVHRR	Mix of orbiting (MODIS) and geostationary (METOP-2 AVHRR). Mostly > 1 km resolution. Daily to 15-minute revisit. For radiation budgets, meteorological information.

Source: GSARS (2017).

Remote sensing data are being used and adapted for countless purposes in farm management, agricultural programs, agricultural statistics, and empirical agricultural economics. GSARS (2017) provides a comprehensive overview of the uses of remote sensing in agricultural statistics, including land cover mapping, the design of sampling frames, crop mapping, crop area and yield estimation, and early warning systems. With Earth Observation data becoming available more frequently and at increasing granularity, recent research has focused on facilitating and validating

the use of these data for different cropping systems, at scale, and in a timely fashion (Defourny et al., 2019). Recent studies have also focused on developing and validating methods and standards for the efficient collecting of *in situ* ground-truthing data for model calibration (Azzari et al., 2021; d’Andrimont, 2018; Paliwal and Jain, 2020).

For empirical applications in agricultural economics, remote sensing data offer the promise of far greater accuracy, objectivity, temporal resolution, and coverage, than could be achieved through traditional survey methods relying on farmers’ self-reporting. However, remote sensing data sets are not immune from measurement error. Michler et al. (2020) identify three sources of error when using remote sensing weather data in conjunction with survey data. Errors can be introduced through the measurement technology, the algorithm to convert the measurement into a variable for analytical use (e.g., rainfall), or the resolution of the data. Errors can also occur in linking remote sensing data to the household, plot, or farm on which the analysis is run, as well as by using variables that are not ‘fit for purpose’ from an agronomic perspective.

The use of remote sensing for crop area estimation and crop and yield mapping is now widespread, with a continuous flow of new competing products being developed by public sector agencies, academics, and the private sector, often in partnership. However, remote sensing presents specific challenges in smallholder systems, which require high resolution and often incorporate inter-cropping patterns that are hard to characterize based on satellite data (Burke and Lobell, 2017; Jain et al., 2016; Jin et al., 2019; Rustowicz et al., 2019). Thus, remote sensing and ground data are much more productively seen as complements rather than substitutes. The use of survey data, particularly objective measures such as crop cuts, for ground-truthing and training models based on satellite data can greatly increase the accuracy of the remote sensing predictions (see Lobell et al., 2019; d’Andrimont, 2018; Paliwal and Jain, 2020 for yield measurement and Hengl et al., 2020 for global soil mapping). The combined use of multiple sources is the most promising avenue for agricultural data systems to minimize error and maximize coverage. As for climate variables, users should be aware of the error structures present in modeled estimates when using them as independent variables in econometric analyses. One key obstacle to using Earth Observation data in conjunction with spatially explicit survey data is that of overcoming confidentiality concerns. For some years now, the United States Department of Agriculture has been aware of the lack of

precise spatial information as a major weakness of their flagship ARMS survey, limiting its value for a range of applications. Other international survey programs such as the Demographic and Health Survey (DHS) and the Living Standards Measurement Study (LSMS) adopt protocols to publicly disseminate ‘masked’ coordinates while preserving anonymity. However, researchers and the global statistical community are still searching for dissemination standards that can maximize the value of spatially explicit data for analytical applications while also preserving anonymity (Croft et al., 2021).

Crowdsourced and citizen-generated data

An innovative source of data that is likely to be increasingly used for research in the coming years is **citizen-generated data** (Lämmerhirt et al., 2018). This includes data generated via **crowdsourcing**, that is, by enlisting a large ‘crowd’ of individuals (volunteers or for pay) or devices (e.g. sensors) to collect and share data. In the cognitive science discipline, one third to one half of the scientific papers in top tier journals are now based on crowdsourced data sets (Stewart et al., 2017). However, at the time of this writing, crowdsourced data are still relatively underused in agricultural economics and are more often employed for operational purposes rather than academic work. In economic research more broadly, the disciplines most likely to use such data are those more amenable to the wholesale enlisting of respondents through dedicated platforms, such as labor market or consumer research. Citizen-generated data are already contributing or demonstrating the potential to contribute to advancing the global data agenda (Fraisl et al., 2020). Their supply and use can be expected to expand rapidly in the coming years, but this will require solutions to overcome issues around quality control and validation (Balázs et al., 2021; Wiggins et al., 2021.)

In the agriculture and food domain, crowdsourced data are more common in price data collection efforts, where agents or volunteers can be recruited to survey markets (UN Global Pulse, 2015; Zeug et al., 2017; Ochieng and Baulch, 2020). They are also used for obtaining climate data, such as rainfall, which is less correlated in space (Minet et al., 2017) and can be crowdsourced by connecting micro rain gauges to the internet (Van de Giesen et al., 2014). Another option is soil data collection, which can be crowdsourced to farmers using smartphone apps to collect soil profile

information (Herrick et al., 2013). One study crowdsourced the visual interpretations of satellite imagery from popular mapping applications to estimate the global distribution of field size (Lesiv et al., 2019). In a review article, Ebitu et al. (2021) identify data collection as the main current thrust for citizen science in agriculture, with key challenges including validation procedures, but primarily the recruitment, motivation, and retention of volunteers.

Citizen-generated data are attractive due to their potential to return data at high levels of spatial and temporal resolution with relatively limited costs. However, these data present significant limitations in their representativeness and the quality of the data generation process that must be understood and managed for statistical inference. Based on a review of survey data, Wiggins et al. (2011) propose a quality assurance framework for citizen science data organized along two categories of sources of errors (which may derive from participants or field protocols) and three entry points in the data production process. While recognizing the huge potential of citizen science data for agriculture and beyond, it is clear that before it can become mainstreamed in data production, more effort must go into ensuring that data collected through “volunteers” with varying levels of expertise and commitment are of acceptable quality (Bonter and Cooper, 2012).

Mehrabi et al. (2021) warn of an emerging global divide in data-driven farming, linked to the differential access to mobile data technologies for low-resourced farmers, particularly in Africa, as a result of a combination of differential ownership of mobile devices, poorer data connection, and connectivity costs. However, the rapid increase in both mobile phone ownership and phone coverage in most countries bodes well for a more widespread adoption of phone data collection. In the cognitive science literature, where crowdsourced data are mainly generated via the Amazon Mechanical Turk platform, concerns have arisen on the professionalization of the individuals contributing the data, with many of them sharing information on internet fora in ways that pose concerns for the independence of the observations (Stewart et al., 2017). Statistics Canada is one of a few statistical offices that have actively published data generated through crowdsourcing, for public policy applications ranging from urban planning to gauging the price of marijuana on the illegal market ahead of its legalization. Tellingly, such data are not accompanied by indications regarding their accuracy (including bias and coverage) that accompany other published statistics (Statistics Canada, 2021).

Methodologies for validating and correcting crowdsourced data through post-stratification efforts (Arbia et al., 2020) or other efforts to assess and improve the bias and variability of the estimates are now starting to emerge (Buil-Gil et al., 2020). With their further development, crowdsourced data will surely become an increasingly important source of data for agricultural economics applications.

Phone surveys

Phone surveys have been around for decades and are in fact part and parcel of the survey data collection in several high-income countries (NRC, 2008; Slavec and Toninelli, 2015). In low-income countries, phone surveys were for some time confined predominantly to the collection of data in conflict- or disaster-affected areas where ground operations are more constrained (Hoogeveen and Pape, 2020), or in urban areas where phone ownership and coverage is higher. However, their adoption has quickly become ubiquitous with the onset of the COVID-19 pandemic in 2020, as statistical offices and practitioners increasingly recognize how phone surveys can become an integral part of a modernized survey system beyond the contingency of the pandemic response period (Glazerman et al., 2020; Young Lives 2020; Josephson et al., 2021).

There are specific coverage concerns for phone surveys linked to the extent and patterns of (mobile) phone penetration, which can be expected to be correlated with variables of interest. Such concerns are far more severe in low-income countries, where phone penetration has been increasing but is still far from universal, and specifically in rural areas, where agricultural economists often focus their research interests (Dillon, 2012; Ballivian et al., 2015; Leo et al., 2015; Lamanna et al., 2019; Mehredi et al., 2021; GSMA, 2020; Dabalen et al., 2016).

During the COVID-19 pandemic, phone surveys allowed for the possibility to contact respondents amid widespread travel and social distancing restrictions, without exposing them or the enumerators to a health risk. Phone surveys can also generate much more frequent data relative to face-to-face interviews, due to their reduced cost and simplified logistics (e.g., not requiring travel). This can limit survey error for variables that are more prone to recall error (such as

agricultural labor (Arthi et al., 2018) or continuous crop production estimates (Kilic et al., 2021), as well as increase the temporal dimension of data collection for outcomes that have low autocorrelation (McKenzie, 2011), or where short-term overtime changes are of the essence, as is the case for the study of resilience (Knippenberg et al., 2019).

Concerns remain for the representativeness and coverage of phone surveys, not only for specific households that may be less likely to have access to a phone connection, but also for individuals who are less likely to be phone owners or are otherwise less represented in phone survey samples (Leo et al., 2015; Brubaker et al., 2021). Such issues can be mitigated when the phone survey sampling frame is based on an adequate set of information on observable household and individual characteristics, as is the case when the phone survey is tied to a recent representative face-to-face survey that collected respondent phone numbers (Ambel et al., 2021). While the sample size of phone surveys using this approach is limited by the sample size of the existing representative survey, phone surveys that use a sampling strategy based on sampling numbers from a list or via Random Digit Dialing (RDD) usually lack sociodemographic information associated with each phone number, making it harder to assess and improve their representativeness (Henderson and Rosenbaum, 2020; Himelein et al., 2020).

Other limitations of phone surveys are related to the type of information that can be asked over the phone, both because of content that respondents may not feel comfortable sharing over the phone, as well as the overall interview length (Abay et al. 2021). Even so, recent experience has demonstrated the value of collecting information over the phone on issues related to agriculture and food security (Amankwah and Gourlay, 2020; Hirvonen et al., 2021), charting the way for a survey research and implementation agenda to leverage the integration of high-frequency data collection via phones and other mobile technology with traditional face-to-face surveys. Such a mixed-mode approach can carry the added advantage of freeing up space in face-to-face surveys from items that can be collected via remote data collection to generate data that are characterized by both reduced survey error and higher temporal resolution. Mixed-mode models can also be instrumental for achieving the temporal resolution needed for many indicators, as well as for providing a low-cost platform to collect more accurate data on high-frequency, repeated occurrences, such as labor allocation in agriculture and other time use data. This is a likely

direction for investment in the survey research agenda in the coming years, where the involvement of agricultural economists in influencing the structure and features of the resulting data will be paramount.

Panel data

Understanding agriculture and the fast transformation processes ongoing in all countries at different stages of development requires panel data. Partly in response to this renewed awareness, we have recently witnessed a surge in the availability of panel data related to agriculture and rural development in low- and middle-income countries. While for decades, the ICRISAT village study (Walker and Ryan, 1990) was one of the few longitudinal data sets allowing research on agricultural and rural livelihoods, over the past two decades the availability of such data sets has increased substantially, even if they remain limited in numbers and geographic coverage. Examples include the panel data set for Ghana collected by researchers at Yale and the Institute of Statistical, Social, and Economic Research in Ghana, the panel data sets collected by statistical offices in eight Sub-Saharan African countries under the World Bank's LSMS-ISA program, the National Income Dynamics Study (NIDS) in South Africa, the Kagera Health and Development Survey in Tanzania, the Family Life Surveys in Indonesia and Mexico, and the panel data collected by IFPRI in several countries in Asia and Africa, by the Tegemeo Institute in Kenya, and by Michigan State in Zambia, among others. These surveys have generated an invaluable wealth of research and contributed to answering key policy questions that cross-sectional data have been unable to convincingly address.

We have discussed above (see section 3.5) several actions that can be taken to manage attrition in panel data, whether ex-ante by improving the design and implementation of tracking protocols, or ex-post. The availability and penetration of mobile phones and the growing adoption of CAPI have been important innovations that have enabled implementing and improving the tracking outcomes for longitudinal surveys in low-income countries. Collecting as many contact numbers as feasible at baseline greatly improves the likelihood of being able to recontact households that move between survey waves, and has also played a fundamental role in allowing the longitudinal tracking of households for phone surveys during the COVID-19 pandemic (Glazerman et al., 2020;

Gourlay et al., 2021). Georeferencing households is an additional technology-based solution that can help relocate the site of the dwelling in areas where these are not otherwise clearly marked or identifiable (Witoelar, 2011).

Additionally, the technology embedded in CAPI applications is providing new approaches for survey designers and implementers to understand and manage attrition (Kreuter, 2013) as well as to improve data quality through better remote supervision. Specifically, the paradata produced during CAPI interviews enables the understanding of certain features that predict attrition as they materialize during the course of the interview, including enumerator effects. These paradata can inform actions to minimize attrition and monitor individuals at higher risk of dropping out of the sample, thus countering the predominant coverage issue for longitudinal data (Mercer, 2012; Roßmann and Gummer, 2016).

Finally, following the onset of the COVID-19 pandemic, the availability of well-established long-term longitudinal studies put countries at an advantage for rapidly shifting to high-frequency phone surveys to monitor the impact of the pandemic. This served to fill critical data demands, while also reducing the potential coverage biases of phone surveys by providing better sampling frames and a wealth of information for the ex-post mitigation of bias.

6. Conclusions

Agricultural data continue to suffer from lack of availability, poor quality, and incomplete coverage. However, in recent years, increasing data demands and emerging policy questions such as climate change and demographic trends, among others, have driven innovation in the sector, with rapid technological change and methodological advances providing an opportunity to collect more and better data at lower cost. In the past two decades, technology has expanded the data production frontier to generate more accurate, granular, and frequent data within shrinking budget envelopes. These innovations have been accompanied by greater attention to issues of measurement error and coverage, focused on ways to attenuate trade-offs and achieve both high

accuracy and high representativeness to the greatest extent possible, and by greater rigor in testing the validity of changes in methods via randomized validation exercises.

This paper is a testament to the increased importance of data and data quality issues within the agricultural economics profession. Researchers hold the power and responsibility to make wiser design choices throughout the data production process. However, reaching the full potential of improvements in data structures for producing policy-relevant empirical analysis may require changes in researchers' incentives and priorities to generate knowledge that is accurate, relevant and credible. For instance, a recent evidence synthesis paper exposes a striking disconnect between empirical agricultural and social science research and policy questions (Porciello et al., 2020).

Throughout the paper, we have highlighted the importance of improving agricultural data structures for empirical analysis, while accounting for the inherent trade-offs involved in designing data collection for agricultural research and policy. Measurement error creates both internal and external validity issues that limit causal inference and descriptive understanding of national agricultural systems. Coverage biases also create internal and external validity issues, particularly when limited coverage biases the testing of underlying mechanisms that drive agricultural choices.

While surveys remain the linchpin of agricultural policy analysis, other traditional data sources such as administrative data and agricultural censuses, as well as new data like Earth Observation and remote sensing data, play equally important roles in improving the coverage of agricultural data in its many domains. Additionally, alternative data sources such as citizen-generated data and methods such as machine learning, while not yet mainstreamed in agricultural data production, offer tremendous opportunities for the future. To achieve their potential, these newer data sources require fully developed quality assurance frameworks to address multiple sources of errors and biases, just as traditional ones do.

As data users become more integrated into data system design, data systems can be better designed for empirical research and policy to minimize measurement error. As emphasized by many authors, non-classical measurement error and its effects vary by sample and are not necessarily adequately

treated and corrected using ex-post econometric tools. Nonetheless, trade-offs are inevitable, as increased coverage can lead to measurement error and internal validity concerns, while low coverage reduces policy relevance and the external validity of parameter estimates.

To promote more systemic learning, validation studies and experimentation must be carried out more systematically within or in parallel to other data collection efforts, and lessons learned from the existing vast body of research in the impact evaluation literature must be streamlined and systematized to offer guidelines on best practices for researchers. Specifically, we propose bridging the gap between the impact evaluation literature and observational studies by methodically incorporating survey experiments to validate new methods and types of data collection. The empirical standard in many validation studies is to use a “gold standard” as *numeraire*, although such “gold standard” metrics are also likely to be measured with error. As a result, many of the available validation studies tend to measure error relative to a standard deemed “closer to the truth”. While technology presents an opportunity to benchmark agricultural measures and generate more objective benchmarks for validation purposes (e.g., DNA fingerprinting to measure improved seed variety), these processes are often considered too costly to be conducted at scale. However, the rapidly decreasing costs and diffusion of new technologies bode well for the future. Furthermore, future survey experiments need to expand the set of econometric techniques to identify unbiased effects of survey design choices beyond pairwise comparisons (Dillon et al., 2019), as has been the case in labor economics, where even the ‘gold standard’ of United States administrative data has been challenged (Abowd and Stinson, 2013).

Fostering greater integration and interoperability across data sources would also allow for more opportunities for minimizing measurement error while maximizing spatial and temporal coverage. As shown, sample surveys have been used to ground-truth remote sensing imagery for the estimation of crop productivity and other agricultural metrics from space. These experiments are examples of how reducing measurement error and improving coverage can simultaneously be achieved through better data interoperability. This is best done when proper design choices are made ex-ante, so as to also minimize the measurement errors of the ground data. Achieving greater reliability of remote sensing data could radically improve the geographic granularity, timeliness, and frequency of agricultural estimates, while also potentially constraining costs. Attaining such a

goal will require the better coordination and acceleration of research efforts, including the production of multi-purpose ground layers of high-quality measurements.

Maximizing coverage of agricultural data also requires improving other traditional sources such as routine data systems and agricultural censuses. The weak data quality of both sources, as well as the low periodicity and predictability of agricultural censuses, particularly in lower-income countries and regions, remain matters of concern. With regards to administrative data, underfunding and the persistent neglect of extension services in past decades are responsible for the current unenviable state of affairs. Digitalization and the adoption of technological solutions can accelerate progress in this area. Furthermore, linking administrative data to newer data sources such as crowdsourced data or high-frequency community surveys through sentinel sites could go a long way towards enhancing the statistical rigor of administrative data. Rethinking administrative data collection and its interoperability with other data sources, while also ensuring better access, should be prioritized to contribute to minimizing error and maximizing coverage of agricultural data. The trend towards greater reliance on administrative data is well advanced in more developed economies, with low- and middle-income countries lagging behind.

New data sources and modes of data collection such as phone or web surveys as well as crowdsourcing and other forms of citizen-generated data offer tremendous potential to improve the availability and frequency of agricultural data. However, to fully exploit these opportunities, better methods are needed to account for likely biases due to selectivity and under-coverage. It is also important to raise awareness, particularly among young researchers, of the pitfalls of ignoring these potential errors and to build their capacity in addressing them, both at the design and analytical stages.

Finally, relying on direct measurements in contrast to the more common practice of asking farmers to self-report, often based on long recalls, has become steadily more feasible due to the declining cost of technology. Nonetheless, cost considerations remain an issue in implementing such methods on a full sample or, in the case of agricultural censuses, on the entire population of concern. Using direct measurement on a sub-sample of households, combined with the rigorous use of imputation methods, may be a more viable way to improve the accuracy of agricultural data.

Bibliography

- Abate, G., Brauw, A., Gibson, J., Hirvonen, K., & Wolle, A. (2020). *Telescoping Causes Overstatement in Recalled Food Consumption: Evidence from a Survey Experiment in Ethiopia*. IFPRI Discussion Paper 1976. Tech. rep., Washington, DC: International Food Policy Research Institute (IFPRI).
- Abay, K. (2020). Measurement Errors in Agricultural Data and their Implications on Marginal Returns to Modern Agricultural Inputs. *Agricultural Economics*, 51. doi:10.1111/agec.12557
- 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. doi:10.1016/j.jdeveco.2019.03
- Abay, K. A., Berhane, G., Hoddinott, J., & Tafere, K. (2021). Assessing Response Fatigue in Phone Surveys: Experimental Evidence on Dietary Diversity in Ethiopia. Policy Research Working Paper. No. 9636. World Bank, Washington, DC.
- Abay, K. A., Bevis, L., & Barrett, C. B. (2020). Measurement Error Mechanisms Matter: Agricultural Intensification with Farmer Misperceptions and Misreporting. *American Journal of Agricultural Economics* 103 (2).
- Aceves-Bueno, E., Adeleye, A., Feraud, M., Huang, Y., Tao, M., Yang, Y., & Anderson, S. (2017). The Accuracy of Citizen Science Data: A Quantitative Review. *Bulletin of the Ecological Society of America*, 98, 278-290. doi:10.1002/bes2.1336
- Akogun, O. B., Dillon, A. S., Friedman, J., Prasann, A., & Serneels, P. M. (2020). Productivity and Health: Physical Activity as a Measure of Effort. *The World Bank Economic Review*.
- Alcser, K., Clemens, J., Holland, L., Guyer, H., & Hu, M. (2016). Interviewer recruitment, selection, and training. *Guidelines for Best Practice in Cross-Cultural Surveys*, 419–468.
- Ali, D. A., & Deininger, K. (2014). *Is there a farm-size productivity relationship in African agriculture ? evidence from Rwanda*. Tech. rep., The World Bank.
- Ali, D. , K. Deininger and A. Harris (2019). Does Large Farm Establishment Create Benefits for Neighboring Smallholders? Evidence from Ethiopia. *Land Economics*. 95 (1)
- 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.
- Amankwah, A. & Gourlay, S. (2021). Impact of COVID-19 Crisis on Agriculture : Evidence from Five Sub-Saharan African Countries (English). LSMS Integrated Surveys on Agriculture Washington, D.C.: World Bank Group.
- Amaya, A., Bach, R., Keusch, F., & Kreuter, F. (2019). New Data Sources in Social Science Research: Things to Know Before Working With Reddit Data. *Social Science Computer Review*, 089443931989330. doi:10.1177/0894439319893305
- Amaya, A., Biemer, P., & Kinyon, D. (2020). Total Error in a Big Data World: Adapting the TSE Framework to Big Data. *Journal of Survey Statistics and Methodology*, 8, 89-119. doi:10.1093/jssam/smz056
- Ambel, A., K. Mcgee and A. Tsegay (2021). *Reducing Bias in Phone Survey Samples: Effectiveness of Reweighting Techniques Using Face-to-Face Surveys as Frames in Four African Countries*. Policy Research Working Paper no. 9676. The World Bank. Washington, DC.

- Ambler, K., Herskowitz, S., & Maredia, M. (2020). *Are we done yet? Response fatigue and rural livelihoods* (Vol. 1980). Intl Food Policy Res Inst.
- Angrist, J. and J.S. Pischke (2010). The Credibility revolution in Empirical Economics. How Better Research Design is Taking the Con out of Econometrics. *Journal of Economic Perspectives*. 24(2)
- Arbia, G., Solano-Hermosilla, G., Micale, F., Nardelli, V., & Genovese, G. (2020). Post-sampling crowdsourced data to allow reliable statistical inference: the case of food price indices in Nigeria. *Post-sampling crowdsourced data to allow reliable statistical inference: the case of food price indices in Nigeria*.
- Arthi, V., Beegle, K., De Weerd, J., & Palacios Lopez, A. (2017). Not your average job: Measuring farm labor in Tanzania. *Journal of Development Economics*, 130. doi:10.1016/j.jdeveco.2017.10.005
- Ashenfelter, O. and A. Krueger (1994). Estimates of the Economic Return to Schooling from a Sample of Twins. *American Economic Review*. 84(5)
- Ashour, M., Gilligan, D., Hoel, J., & Karachiwalla, N. (2018). Do Beliefs About Herbicide Quality Correspond with Actual Quality in Local Markets? Evidence from Uganda. *The Journal of Development Studies*, 55, 1-22. doi:10.1080/00220388.2018.1464143
- 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.
- Bachewe, F. N., Minten, B., Taffesse, A. S., Pauw, K., Cameron, A., & Endaylalu, T. G. (2020). Farmers' grain storage and losses in Ethiopia: Measures and associates. *Journal of Agricultural and Food Industrial Organization*, 18. doi:10.1515/jafio-2019-0059
- Bailey, D., Trotter, M., Knight, C., & Thomas, M. (2018). Use of GPS tracking collars and accelerometers for rangeland livestock production research. *Translational Animal Science*, 2. doi:10.1093/tas/ttx006
- Balázs B., Mooney P., Nováková E., Bastin L., Jokar Arsanjani J. (2021) Data Quality in Citizen Science. In: Vohland K. et al. (eds) *The Science of Citizen Science*. Springer, Cham. https://doi.org/10.1007/978-3-030-58278-4_8
- Bakker, B. F., Van Rooijen, J., & Van Toor, L. (2014). The system of social statistical datasets of Statistics Netherlands: an integral approach to the production of register-based social statistics. *Statistical journal of the United Nations ECE*, 30, 411–424. doi:10.3233/SJI-140803
- Ballivian, A., Azevedo, J. P., Durbin, W., & Bank, W. (2015). Using Mobile Phones for High-Frequency Data Collection. In *Mobile Research Methods: Opportunities and Challenges of Mobile Research Methodologies*, 21–39.
- Bardasi, E., Sabarwal, S., & Terrell, K. (2011). How do female entrepreneurs perform? Evidence from three developing regions. *Small Business Economics*, 37, 417-441. doi:10.1007/s11187-011-9374-z
- Bardhan, P., & Udry, C. (1999). *Development microeconomics*. OUP Oxford.
- Barrett, C. B., Gebru, G., McPeak, J. G., Mude, A. G., Vanderpuye-Orgle, J., & Yirbecho, A. T. (2008). Codebook for data collected under the improving pastoral risk management on East African Rangelands (PARIMA) Project. *Unpublished, Cornell University*.
- Barrett, C., T. Reardon, J. Swinnen and D. Zilberman (forthcoming). Agro-Food Value Chain Revolutions in Low- and Middle-Income Countries. *Journal of Economic Literature*.

- Baulch, B., Ochieng, D. O., & others. (2020). *Most Malawian maize and soybean farmers sell below official minimum farmgate prices*. Tech. rep., International Food Policy Research Institute (IFPRI).
- Beaman, L., & Dillon, A. (2012). Do household definitions matter in survey design? Results from a randomized survey experiment in Mali. *Journal of Development Economics*, 98, 124-135. doi:10.1016/j.jdeveco.2011.06
- Beaman, L., Benyishay, A., Magruder, J., & Mobarak, A. (2018). Can Network Theory-Based Targeting Increase Technology Adoption? *SSRN Electronic Journal*. doi:10.2139/ssrn.3225815
- Beegle, K., De Weerd, J., & Dercon, S. (2011). Migration and economic mobility in Tanzania: Evidence from a tracking survey. *Review of Economics and Statistics*, 93, 1010–1033.
- Beegle, K., Himelein, K., & Ravallion, M. (2012). Frame-of-reference bias in subjective welfare. *Journal of Economic Behavior & Organization - J ECON BEHAV ORGAN*, 81. doi:10.1016/j.jebo.2011.07.020
- Beegle, K., Olinto, P., Sobrado, C., & Uematsu, H. (2013). The State of the Poor: Where Are The Poor, Where Is Extreme Poverty Harder to End, and What Is the Current Profile of the World's Poor? *World Bank Economic Premise*.
- Bellemare, M. G., Dabney, W., & Munos, R. (2017). A distributional perspective on reinforcement learning. *International Conference on Machine Learning*, (pp. 449–458).
- Belli, Robert F., et al. "Calendar and question-list survey methods: Association between interviewer behaviors and data quality." *Journal of Official Statistics* 20.2 (2004): 185.
- Benami, E., Jin, Z., Carter, M. R., Ghosh, A., Hijmans, R. J., Hobbs, A., . . . Lobell, D. B. (2021). Uniting remote sensing, crop modelling and economics for agricultural risk management. *Nature Reviews Earth & Environment*, 1–20.
- 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. doi:https://doi.org/10.1016/j.worlddev.2017.12.009
- Beullens, K., & Loosveldt, G. (2016). Interviewer Effects in the European Social Survey. In *Survey Research Methods* (Vol. 10, No. 2, Pp. 103-118). European Survey Research Association.
- Bevis, L., & Barrett, C. (2019). Close to the edge: High productivity at plot peripheries and the inverse size-productivity relationship. *Journal of Development Economics*, 143, 102377. doi:10.1016/j.jdeveco.2019.102377
- Biagas, D., E. Abayomi, J. Rodhouse and H. Ridolfi (2019). Examining Interviewer Effects on the Agricultural Labor Survey: A Mixed-Methods Approach. 2019 *Workshop: Interviewers and Their Effects from a Total Survey Error Perspective*. 34. <http://digitalcommons.unl.edu/sociw/34>
- Biemer, P. (2009). Chapter 12 - Measurement Errors in Sample Surveys. In C. R. Rao (Ed.), *Handbook of Statistics* (Vol. 29, pp. 281-315). Elsevier. doi:https://doi.org/10.1016/S0169-7161(08)00012-6
- Biemer, P. (2010). Total Survey Error: Design, Implementation, and Evaluation. *The Public Opinion Quarterly*, 74, 817-848. doi:10.2307/40985407
- Biemer, P. (2017). Errors and inference. *Big data and social science: A practical guide to methods and tools*, 265–298.

- Biemer, P. P., Groves, R. M., Lyberg, L. E., Mathiowetz, N. A., & Sudman, S. (1991). *Measurement errors in surveys*. John Wiley & Sons.
- Biemer, P., & Lyberg, L. (2003). *Introduction to Survey Quality* (Vol. 335). doi:10.1002/0471458740
- Blaydes, Lisa, and Rachel M. Gillum. "Religiosity-of-interviewer effects: Assessing the impact of veiled enumerators on survey response in Egypt." *Politics & Religion* 6.3 (2013): 459-482.
- Bold, T., Kaizzi, K., Svensson, J., & Yanagizawa-Drott, D. (2017). Lemon Technologies and Adoption: Measurement, Theory and Evidence from Agricultural Markets in Uganda*. *Quarterly Journal of Economics*, 132. doi:10.1093/qje/qjx009
- Bonter, D.N. and C.B. Cooper (2012). Data Validation in Citizen Science: a Casa Study from Project FeederWatch. *Frontiers in Ecology and the Environment*. 10 (6).
- Bound, J., Brown, C., & Mathiowetz, N. (2001). Chapter 59 - Measurement Error in Survey Data. In *Handbook of Econometrics*, Volume 5, J. J. Heckman, & E. Leamer (Eds.). Elsevier. doi:https://doi.org/10.1016/S1573-4412(01)05012-7
- Brubaker, J.M., T. Kilic and P. Wollburg (2021). *Reopresentativeness of Individual-Level Data in Covid-19 Phone Surveys: Findings from Sub-Saharan Africa*. Policy Research Working Paper 9660. World Bank, Washington, DC.
- Buil-Gil, D., Solymosi, R., & Moretti, A. (2020). Nonparametric Bootstrap and Small Area Estimation to Mitigate Bias in Crowdsourced Data. In T. D. C.A. Hill, & L. E. Lyberg (Eds.), *Big Data Meets Survey Science* (pp. 487-517). John Wiley & Sons, Ltd. doi:https://doi.org/10.1002/9781118976357.ch16
- Burke, W., Frossard, E., Kabwe, S., & Jayne, T. (2019). Understanding fertilizer adoption and effectiveness on maize in Zambia. *Food Policy*. doi:10.1016/j.foodpol.2019.05.004
- Caeyers, B., Chalmers, N., & De Weerd, J. (2010). A comparison of CAPI and PAPI through a randomized field experiment. *SSRN Electronic Journal*. doi:10.2139/ssrn.1756224
- Cannell, C. F., Marquis, K. H., & Laurent, A. (1976). *A summary of studies of interviewing methodology* (Vol. 77). Department of Health, Education, and Welfare, Public Health Service, Health
- Carfagna, E., & Gallego, F. J. (2005). Using Remote Sensing for Agricultural Statistics. *International Statistical Review*, 73, 389-404. doi:https://doi.org/10.1111/j.1751-5823.2005.tb00155.x
- Carletto, C., Aynekulu, E., Gourlay, S., & Shepherd, K. (2017a). *Collecting the dirt on soils: advancements in plot-level soil testing and implications for agricultural statistics*. The World Bank.
- Carletto, C., Corral, P., & Guelfi, A. (2017b). Agricultural commercialization and nutrition revisited: Empirical evidence from three African countries. *Food Policy*, 67, 106-118. doi:https://doi.org/10.1016/j.foodpol.2016.09.020
- Carletto, C., Deininger, K., Savastano, S., & Muwonge, J. (2012). Using diaries to improve crop production statistics: Evidence from Uganda. *Journal of Development Economics*, 98(1), 42-50.
- Carletto, C., Gourlay, S., Murray, S., & Zezza, A. (2017c). Cheaper, faster, and more than good enough: Is GPS the new gold standard in land area measurement? *Survey Research Methods*, 11, pp. 235–265.
- Carletto, G., Gourlay, S., Murray, S., & Zezza, A. (2016). *Land Area Measurement in Household Surveys: A Guidebook*. Tech. rep., Washington DC: World Bank.

- Carletto, G., Ruel, M., Winters, P., & Zezza, A. (2015). Farm-Level Pathways to Improved Nutritional Status: Introduction to the Special Issue. *The Journal of Development Studies*, 51, 945-957. doi:10.1080/00220388.2015.1018908
- Carletto, G., Zezza, A., & Banerjee, R. (2013). Towards Better Measurement Of Household Food Security: Harmonizing Indicators And The Role Of Household Surveys. *Global Food Security*, 2, 30-40. doi:10.1016/j.gfs.2012.11.006
- Carroll, R., Ruppert, D., Stefanski, L., & Crainiceanu, C. (2006). *Measurement Error in Nonlinear Models: A Modern Perspective*. Boca Raton: Chapman & Hall.
- Carter, R., Lau, M., Johnson, V., & Kirkinis, K. (2017). Racial Discrimination and Health Outcomes Among Racial/Ethnic Minorities: A Meta-Analytic Review. *Journal of Multicultural Counseling and Development*, 45, 232-259. doi:10.1002/jmcd.12076
- Chambers, R. G. (1988). *Applied production economics: a dual approach*.
- Chambers, R. G., & Quiggin, J. (2000). *Uncertainty, production, choice, and agency: the state-contingent approach*. Cambridge University Press.
- Chamoso, P., Raveane, W., Parra, V., & González, A. (2014). UAVs Applied to the Counting and Monitoring of Animals. *Advances in Intelligent Systems and Computing*, 291. doi:10.1007/978-3-319-07596-9_8
- Chesher, A., & Schluter, C. (2002). Welfare Measurement and Measurement Error. *The Review of Economic Studies*, 69, 357-378.
- Cohen, A. (2019). Estimating Farm Production Parameters with Measurement Error in Land Area. *Economic Development and Cultural Change*, 68, 305-334. doi:10.1086/700557
- Dabalen, A., Etang, A., Hoogeveen, J., Mushi, E., Schipper, Y., & Engelhardt, J. von. (2016). *Mobile Phone Panel Surveys in Developing Countries: A Practical Guide for Microdata Collection*. The World Bank.
- Da Re, D., Gilbert, M., Chaiban, C., Bourguignon, P., Thanapongtharm, W., Robinson, T., & Vanwambeke, S. (2020). Downscaling livestock census data using multivariate predictive models: Sensitivity to modifiable areal unit problem. *PLOS ONE*, 15, e0221070. doi:10.1371/journal.pone.0221070
- Das, J., J. Hammer and C. Sanchez-Paramo (2012). The Impact of Recall Periods on reported Morbidity and Health Seeking Behavior. *Journal of Development Economics*. 98
- Das, N., Davies, E., Dillon, A., Glazerman, S., & Rosenbaum, M. (2021). Optimal Timing for Random Digit Dialing. *Global Poverty Research Lab Working Paper* No. 21-107.
- Datashift. (2015). *What is Citizen-Generated Data And What Is The DataShift Doing To Promote It?*
- Davis, D. W., Silver, B. D. (2003). "Stereotype Threat and Race of Interviewer Effects in a Survey on Political Knowledge," *American Journal of Political Science*, 47, 33-45.
- Davis, R. E., Couper, M. P., Janz, N. K., Caldwell, C. H., Resnicow, K. (2010). "Interviewer Effects in Public Health Surveys," *Health Education Research*, 25, 14-26.
- De Haan, W., Van Berkel, S., Van Der Asdonk, S., Finkenauer, C., Forder, C., Van Ijzendoorn, M., . . . Alink, L. (2019). Out-of-home placement decisions: How individual characteristics of professionals are reflected in deciding about child protection cases. *Developmental Child Welfare*, 1, 251610321988797. doi:10.1177/2516103219887974
- De Leeuw, E. (2004). To Mix or Not to Mix Data Collection Modes in Surveys. *J Off Stat*, 21.
- De Leeuw, E. D., & Van der Zouwen, J. (1988). Data quality in telephone and face to face surveys: a comparative meta-analysis. *Telephone survey methodology*.

- De Mel, S., McKenzie, D., & Woodruff, C. (2008). Returns to Capital in Microenterprises: Evidence from a Field Experiment. *The Quarterly Journal of Economics*, 123, 1329-1372. doi:10.1162/qjec.2008.123.4.1329
- De Nicola, F and X. Giné (2014). How accurate are Recall Data? Evidence from Coastal India. *Journal of Development Economics*, 106 (1).
- De Weerd, J. K. Beegle; J. Friedman and J. Gibson (2016). The Challenges of Measuring Hunger through Survey. *Economic Development and Cultural Change*. 64(4).
- De Weerd, J., J. Gibson and K. Beegle (2020). What Can We Learn from Experimenting with Survey Methods? *Annual Review of Resource Economics*. 12.
- Deaton, A., & Zaidi, S. (2002). *Guidelines for constructing consumption aggregates for welfare analysis* (Vol. 135). World Bank.
- Deininger, K., Byerlee, D., Lindsay, J., Norton, A., Selod, H., & Stickler, M. (2011). *Rising Global Interest in Farmland: Can it Yield Sustainable and Equitable Benefits?* The World Bank.
- Dell, M., Jones, B. F., & Olken, B. A. (2014). What Do We Learn from the Weather? The New Climate–Economy Literature. *Journal of Economic Literature*, 52, 740–798.
- Deming, E. W. (2006). On Errors in Surveys (An Excerpt). *The American Statistician*, 60, 34-38. doi:10.1198/000313006X91755
- Deming, W. E. (1944). On errors in surveys. *American Sociological Review*, 9, 359–369.
- Desiere, S., & Jolliffe, D. (2018). Land productivity and plot size: Is measurement error driving the inverse relationship? *Journal of Development Economics*, 130, 84-98. doi:https://doi.org/10.1016/j.jdeveco.2017.10.002
- Di Falco, S., Veronesi, M., & Yesuf, M. (2011). Does Adaptation to Climate Change Provide Food Security? A Micro-Perspective from Ethiopia. *American Journal of Agricultural Economics*, 93, 825-842. doi:10.1093/ajae/aar006
- Di Maio, M., & Fiala, N. (2019). Be Wary of Those Who Ask: A Randomized Experiment on the Size and Determinants of the Enumerator Effect. *The World Bank Economic Review*, 34, 654-669. doi:10.1093/wber/lhy024
- Diego-Rosell, P., Nichols, S., Srinivasan, R., & Dilday, B. (2020). Assessing Community Wellbeing Using Google Street-View and Satellite Imagery. In T. D. C.A. Hill, & L. E. Lyberg (Eds.), *Big Data Meets Survey Science* (pp. 435-486). John Wiley & Sons, Ltd. doi:https://doi.org/10.1002/9781118976357.ch15
- Dillon, A., & Mensah, E. (2020). Respondent Biases in Household Surveys. *Global Poverty Lab Working Paper*.
- Dillon, A., & Rao, L. (2018). Land Measurement Bias: Comparisons from Global Positioning System, Self-Reports, and Satellite Data. *SSRN Electronic Journal*. doi:10.2139/ssrn.3188522
- Dillon, A., Carletto, G., Gurlay, S., Wollburg, P., & Zezza, A. (2021a). *Advancements in Data Collection Methods for Agricultural Surveys: Lessons from the LSMS-ISA and Beyond*. Tech. rep., FAO.
- Dillon, A., Glazerman, S., & Rosenbaum, M. (2021b). Understanding Response Rates in Random Digit Dial Surveys. *Global Poverty Research Lab Working Paper No. 21-105*.
- Dillon, A., Glazerman, S., and Rosenbaum, M. (2021c). Messaging to Improve Response Rates: Effectiveness of Pre-Survey SMS Messages. *Global Poverty Research Lab Working Paper No. 21-106*.

- Dillon, A., Gourlay, S., Mcgee, K., & Oseni, G. (2019). Land Measurement Bias and Its Empirical Implications: Evidence from a Validation Exercise. *Economic Development and Cultural Change*, 67. doi:10.1086/698309
- Dillon, A., Karlan, D., Udry, C., & Zinman, J. (2020). Good identification, meet good data. *World Development*, 127, 104796. doi:10.1016/j.worlddev.2019.104796
- Dillon, B. (2012). Field Report Using Mobile Phones to Collect Panel Data in Developing Countries. *Journal of International Development*, 527(February 2011), 518–527. <https://doi.org/10.1002/jid>
- DiNardo, J., McCrary, J., & Sanbonmatsu, L. (2006). Constructive proposals for dealing with attrition: An empirical example. *NBER working paper*, 1–46.
- Dinku, T. (2019). Chapter 7 - Challenges with availability and quality of climate data in Africa. In A. M. Melesse, W. Abtew, & G. Senay (Eds.), *Extreme Hydrology and Climate Variability* (pp. 71-80). Elsevier. doi:<https://doi.org/10.1016/B978-0-12-815998-9.00007-5>
- Dobardzic, S., Dengel, C. G., Gomes, A. M., Hansen, J., Bernardi, M., Fujisawa, M., . . . others. (2019). 2019 State of Climate Services: Agriculture and Food Security. *World Meteorological Organization*.
- D'Orazio, M. (2020). *Finding ways of reconciling the choice of frames and maximizing coverage by linking multiple frames has been the subject of recent research under the 50x2030 Data Smart Agriculture initiative*. Tech. rep., FAO.
- Doss, B., Roddy, M., Nowlan, K., Rothman, K., & Christensen, A. (2018). Maintenance of Gains in Relationship and Individual Functioning Following the Online OurRelationship Program. *Behavior Therapy*, 50. doi:10.1016/j.beth.2018.03.011
- Doss, C., & Kieran, C. (2014). Standards for Collecting Sex-Disaggregated Data for Gender Analysis; a Guide for CGIAR Researchers.
- Doss, C., & Quisumbing, A. (2019). Understanding rural household behavior: Beyond Boserup and Becker. *Agricultural Economics*, 51. doi:10.1111/agec.12540
- Doss, C., Kieran, C., & Kilic, T. (2020). Measuring Ownership, Control, and Use of Assets. *Feminist Economics*, 26, 144-168.
- Doss, C., Kovarik, C., Peterman, A., Quisumbing, A., & van den Bold, M. (2015). Gender inequalities in ownership and control of land in Africa: myth and reality. *Agricultural Economics*, 46, 403-434. doi:<https://doi.org/10.1111/agec.12171>
- Eisenhower, D., Mathiowetz, N. A., & Morganstein, D. (1991). Recall Error: Sources and Bias Reduction Techniques. In P. P. Biemer, R. M. Groves, L. E. Lyberg, N. A. Mathiowetz, & S. Sudman (Eds.), *Measurement Errors in Surveys* (pp. 125-144). John Wiley & Sons, Ltd. doi:<https://doi.org/10.1002/9781118150382.ch8>
- Engle-Stone, R., Sununtnasuk, C., & Fiedler, J. L. (2017). Investigating the significance of the data collection period of household consumption and expenditures surveys for food and nutrition policymaking: Analysis of the 2010 Bangladesh household income and expenditure survey. *Food Policy*, 72, 72-80. doi:10.1016/j.foodpol.2017.08.014
- Fafchamps, M. (1993). Sequential Labor Decisions Under Uncertainty: An Estimable Household Model of West-African Farmers. *Econometrica*, 61, 1173–1197.
- Falaris, E. (2003). The effect of survey attrition in longitudinal surveys: evidence from Peru, Cote d'Ivoire and Vietnam. *Journal of Development Economics*, 70, 133-157.

- Falorsi, P.D. and D. Bako (2016). Indirect Sampling, a Way to Overcome the Weakness of the Lists in Agricultural Surveys. Seventh International Conference of Agricultural Statisticians, Rome, Italy. doi: 10.1481/icasVII.2016.f35e
- FAO. (1992). *Collecting Data on Livestock* (Vol. 4). FAO Statistical Development Series.
- FAO. (2002). *Land Tenure and Rural Development*.
- FAO. (2015). World Programme for the Census of Agriculture 2020, Volume 1. *Programme, concepts and definitions*. FAO, Rome, Italy.
- FAO (2015). Handbook on Master Sampling Frames for Agricultural Statistics - Frame Development, Sample Design and Estimation. FAO, Rome, Italy.
- FAO. (2017). *Global database of GHG emissions related to feed crops: Methodology. Version 1*. Livestock Environmental Assessment and Performance Partnership. FAO, Rome, Italy.
- FAO. (2019). *The state of Food and Agriculture 2019. Moving forward on food loss and waste reduction*. Rome, Italy.
- 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: World Bank. FAO, World Bank, and UN Habitat.
- Fermont, A., & Benson, T. (2011). Estimating yield of food crops grown by smallholder farmers: A review in the Uganda context. *International Food Policy Research Institute discussion paper, 01097*.
- Fisher, R. (1926). The Arrangement of Field Experiments. *Journal of the Ministry of Agriculture*, 33, 503-515.
- Flores-Macias F., Lawson C. (2008). "Effects of Interviewer Gender on Survey Responses: Findings from a Household Survey in Mexico," *International Journal of Public Opinion Research*, 20, 100–110.
- Floro, V., Labarta, R., Becerra Lopez-Lavalle, L., Martínez, J., & Ovalle, T. (2017). Household Determinants of the Adoption of Improved Cassava Varieties using DNA Fingerprinting to Identify Varieties in Farmer Fields: A Case Study in Colombia. *Journal of Agricultural Economics*, 69. doi:10.1111/1477-9552.12247
- FLW Protocol Steering Committee. (2016). Food Loss and Waste Accounting and Reporting Standard.
- Foster, A. D., & Rosenzweig, M. R. (2010). Microeconomics of Technology Adoption. *Annual Review of Economics*, 2, 395-424. doi:10.1146/annurev.economics.102308.124433
- Fowler Jr, F. J. (1995). *Improving survey questions: Design and evaluation*. Sage.
- Fowler Jr, F. J. (2004). Reducing interviewer-related error through interviewer training, supervision, and other means. *Measurement errors in surveys*, 259–278.
- Fowler, F. J., & Mangione, T. W. (1985). *The value of interviewer training and supervision*. Center for Survey Research.
- Fraisl, D., Campbell, J., See, L., Wehn, U., Wardlaw, J., Gold, M., Moorthy, I., Arias, R., Piera, J., Oliver, J.L., Masó, J., Penker, M., & Fritz, S. (2020.) Mapping citizen science contributions to the UN sustainable development goals. *Sustainability Science*. 15, 1735–1751. <https://doi.org/10.1007/s11625-020-00833-7>
- Gaddis, I., Oseni, G., Palacios-Lopez, A., & Pieters, J. (2020). Measuring Farm Labor: Survey Experimental Evidence from Ghana. *The World Bank Economic Review*. doi:10.1093/wber/lhaa012
- Gallup. (2012). Listening to LAC (L2L). Washington, DC: World Bank.

- Gennari, P., P. D. Falorsi and C.A. Khalil (2013). The Indirect Sampling as a General Approach for Defining Unbiased Sampling Strategies for integrated Agricultural Surveys. Proceedings of the 59th ISI World Statistics Congress, 25-30 August 2013, Hong Kong.
- Gibson, D. G., Wosu, A. C., Pariyo, G. W., Ahmed, S., Ali, J., Labrique, A. B., . . . Hyder, A. A. (2019). Effect of airtime incentives on response and cooperation rates in non-communicable disease interactive voice response surveys: randomised controlled trials in Bangladesh and Uganda. *BMJ global health*, 4, e001604.
- Gibson, D., Ochieng, B., Kagucia, E., Were, J., Hayford, K., Moulton, L., . . . Feikin, D. (2017). Mobile phone-delivered reminders and incentives to improve childhood immunisation coverage and timeliness in Kenya (M-SIMU): a cluster randomised controlled trial. *The Lancet Global Health*, 5, e428-e438. doi:10.1016/S2214-109X(17)30072-4.
- Gideon, L. (2012). *Handbook of Survey Methodology for the Social Sciences*.
- Glazerman, S., Rosenbaum, M., Sandino, R., & Shaughnessy, L. (2020). *Remote Surveying in a Pandemic: Handbook*. 1–19. <https://www.poverty-action.org/sites/default/files/publications/IPA-Phone-Surveying-in-a-Pandemic-Handbook-Updated-December-2020.pdf>
- Gollin, D., & Udry, C. (2021). Heterogeneity, Measurement Error, and Misallocation: Evidence from African Agriculture. *Journal of Political Economy*, 129, 1-80. doi:10.1086/711369
- Gollin, D., Lagakos, D., & Waugh, M. (2014). The Agricultural Productivity Gap. *American Economic Review: Papers & Proceedings 2014*, 104, 165-170.
- Gonzalez Villalobos, A., and W. H. Wigton (2011). *On Applying Area and Multiple Frame Sampling Methods in a Wide Range of Baseline Agricultural and Rural Survey Programmes*. Proceedings of the 58th World Statistical Congress, Dublin.
- Gottschalk, P., & Huynh, M. (2010). Are Earnings Inequality and Mobility Overstated? The Impact of Nonclassical Measurement Error. *The Review of Economics and Statistics*, 92, 302-315.
- Gourlay, S., Kilic, T., & Lobell, D. (2017). Could the Debate Be Over? Errors in Farmer-Reported Production and Their Implications for the Inverse Scale-Productivity Relationship in Uganda. *Policy Research Working Paper No. 8192*. World Bank, Washington, DC.
- 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. doi:10.1016/j.jdeveco.2019.102376
- Gourlay, S., Kilic, T., Martuscelli, A., Wollburg, P., & Zezza, A. (2021). *High-Frequency Phone Surveys on COVID-19: Best Practices, Open Questions*. Tech. rep., World Bank.
- Gourley, J., Flamig, Z., Vergara, H., Kirstetter, P.-E., Clark, R., Argyle, E., . . . Howard, K. (2017). The FLASH project: improving the tools for flash flood monitoring and prediction across the United States. *Bulletin of the American Meteorological Society*, 98.
- Greenleaf, A. R., Gadiaga, A., Guiella, G., Turke, S., Battle, N., Ahmed, S., & Moreau, C. (2020). Comparability of modern contraceptive use estimates between a face-to-face survey and a cellphone survey among women in Burkina Faso. *PLOS ONE*, 15, 1-15. doi:10.1371/journal.pone.0231819
- Griliches, Z. (1986). *Productivity, R&D, and Basic Research at the Firm Level in the 1970s*. Working Paper, National Bureau of Economic Research. doi:10.3386/w1547

- Grosh, M. and P. Glewwe (2000). *Designing Household Survey Questionnaires for Developing Countries : Lessons from 15 Years of the Living Standards Measurement Study*. Washington, DC: World Bank
- Groves, R. M. (1989). *Survey Errors and Survey Costs*. Wiley.
- Groves, R., & Lyberg, L. (2010). Total Survey Error: Past, Present, and Future. *The Public Opinion Quarterly*, 74, 849-879. doi:10.2307/40985408
- GSARS. (2016). *Guidelines for the Enumeration of Nomadic and Semi-Nomadic (Transhumant) Livestock*.
- GSARS. (2017). *Handbook on Remote Sensing for Agricultural Statistics. Publication prepared in the framework of the Global Strategy to improve Agricultural and Rural Statistics*.
- GSARS. (2018). *Handbook on crop statistics: improving methods for measuring crop area, production and yield. Publication prepared in the framework of the Global Strategy to improve Agricultural and Rural Statistics*.
- GSMA. (2020). Mobile Economy. *GSMA Intelligence*.
- Hale, Robert (1999). Appropriate Role of Remote Sensing in U.S. Agricultural Statistics. FAO regional project - Improvement of Agricultural Statistics in Asia and Pacific Countries (GCP/RAS/171/JPN). FAO, Bangkok, Thailand.
- Hausmann, J. (2001). Mismeasured Variables in Econometric Analysis: Problems from the Right and Problems from the Left. *Journal of Economic Perspectives*, 15, 57-67. doi:10.1257/jep.15.4.57
- Henderson, S., Rosenbaum, M., 2020. Remote Surveying in a Pandemic: Research Synthesis. Innovation for Poverty Action.
- Hengl, T., Heuvelink, G. B., Kempen, B., Leenaars, J. G., Walsh, M. G., Shepherd, K. D., . . . others. (2015). Mapping soil properties of Africa at 250 m resolution: Random forests significantly improve current predictions. *PloS one*, 10, e0125814.
- Hengl, T., Miller, M., Križan, J., Shepherd, K., Sila, A., Kilibarda, M., . . . Crouch, J. (2020). African Soil Properties and Nutrients Mapped at 30–m Spatial Resolution using Two-scale Ensemble Machine Learning. doi:10.21203/rs.3.rs-120359/v1
- Herrick, J., Sala, O., & Karl, J. (2013). Land degradation and climate change: a sin of omission? *Frontiers in Ecology and the Environment*, 11. doi:10.2307/23470470
- Hicks, J. H., Kleemans, M., Li, N. Y., & Miguel, E. (2017). *Reevaluating Agricultural Productivity Gaps with Longitudinal Microdata*. Working Paper, National Bureau of Economic Research. doi:10.3386/w23253
- Hill, C., Biemer, P., Buskirk, T., Callegaro, M., Cazar, A., Eck, A., . . . Sturgis, P. (2019). Exploring New Statistical Frontiers at the Intersection of Survey Science and Big Data: Convergence at “BigSurv18”. *Survey Research Methods*, 13. doi:10.18148/srm/2019.v1i1.7467
- Himelein, K. (2015). Interviewer Effects in Subjective Survey Questions: Evidence From Timor-Leste. *International Journal of Public Opinion Research*, 28, 511-533. doi:10.1093/ijpor/edv031
- Himelein, K., Eckman, S., Kastelic, J., McGee, K., Wild, M., Yoshida, N., Hoogeveen, J. (2020). High Frequency Mobile Phone Surveys of Households to Assess the Impacts of COVID-19. Guidelines on Sampling Design. World Bank, Washington D.C.
- Himelein, K., Eckman, S., & Murray, S. (2014). Sampling Nomads: A New Technique for Remote, Hard-to-Reach, and Mobile Populations. *Journal of Official Statistics*, 30. doi:10.2478/jos-2014-0013

- Hirvonen, K., de Brauw, A., & Abate, G. T. (2021). Food Consumption and Food Security during the COVID-19 Pandemic in Addis Ababa. *American Journal of Agricultural Economics*, 103(3), 772–789. <https://doi.org/10.1111/ajae.12206>
- Hogset, H. and C. B. Barrett (2010). Social Learning, Social Influence, and Projection Bias: a Caution of Inferences Based on Proxy Reporting of Peer Behavior. *Economic Development and Cultural Change*. 58(3).
- Holden, S., Ali, D., Deininger, K., & Hilhorst, T. (2016). A Land Tenure Module for LSMS.
- Hu, Y., & Schennach, S. M. (2008). Instrumental Variable Treatment of Nonclassical Measurement Error Models. *Econometrica*, 76, 195–216.
- Hyslop, R., & Imbens, G. W. (2001). Bias from Classical and Other Forms of Measurement Error. *Journal of Business & Economic Statistics*, 19, 475–481.
- Iarossi, G. (2006). *The Power of Survey Design : A User's Guide for Managing Surveys, Interpreting Results, and Influencing Respondents*. Washington, DC: World Bank.
- ILCA. (1990). *Livestock System Research. Vol. 1, International Livestock Center for Africa, Addis Ababa*.
- Jaleta, M., Tesfaye, K., Kilian, A., Yirga, C., Habte, E., Beyene, H., . . . Erenstein, O. (2020). Misidentification by farmers of the crop varieties they grow: Lessons from DNA fingerprinting of wheat in Ethiopia. *Plos one*, 15, e0235484.
- Japac, L., Kreuter, F., Berg, M., Biemer, P., Decker, P., Lampe, C., . . . Usher, A. (2015). Big Data in Survey Research. *Public Opinion Quarterly*, 79, 839-880. doi:10.1093/poq/nfv039
- Jayne, T.S., M. Muyanga, A. Wineman, H. Ghebru, C. Stevens, M. Stickler, A. Chapoto, W. Anseeuw, D. van der Westhuizen and D. Nyange (2019). Are medium-Scale Farms Driving Agricultural Transformation in Sub-Saharan Africa? *Agricultural Economics*. 50.
- Jensen, N., & Barrett, C. (2017). Agricultural Index Insurance for Development. *Applied Economic Perspectives and Policy*, 39, 199-219. doi:<https://doi.org/10.1093/aep/ppw022>
- Jerven, M. and D. Johnston (2015). Statistical Tragedy in Africa? Evaluating the Database for African Economic Development. *Journal of development Studies*. 51(2).
- Josephson, A., Kilic, T. & Michler, J.D. (2021.) Socioeconomic impacts of COVID-19 in low-income countries. *Nature Human Behavior* 5, 557–565. <https://doi.org/10.1038/s41562-021-01096-7>
- Juran, J. M., & Gryna, F. M. (1980). *Quality planning and analysis*. McGraw-Hill: New York.
- Kasprzyk, D. (2005). Chapter IX. *Measurement error in household surveys: sources and measurement*. in Household Surveys in Developing and Transition Countries, United Nations.
- Kastelic, K. H., Eckman, S., Kastelic, J. G., McGee, K. R., Wild, M., Yoshida, N., & Hoogeveen, J. G. (2020). *High Frequency Mobile Phone Surveys of Households to Assess the Impacts of COVID-19 (Vol. 2): Guidelines on Sampling Design*. Tech. rep., Washington, D.C.: World Bank.
- Kilic, T., & Moylan, H. (2016). *Methodological Experiment on Measuring Asset Ownership from a Gender Perspective*. Tech. rep., World Bank, Washington, DC. © World Bank. doi:10.1596/33653
- Kilic, T., Djima, I., and Carletto, G. (2017). *Mission Impossible? Exploring the Promise of Multiple Imputation for Predicting Missing GPS-Based Land Area Measures in Household Surveys*. doi:10.1596/1813-9450-8138

- Kilic, T., Moylan, H. G., and Koolwal, G. B. (2020b). Getting the (Gender-Disaggregated) Lay of the Land : Impact of Survey Respondent Selection on Measuring Land Ownership and Rights. *Policy Research Working Paper Series*. Washington, DC: World Bank.
- Kilic, T., Moylan, H., Ilukor, J., Mtengula, C., & Pangapanga-Phiri, I. (2021). Root for the tubers: Extended-harvest crop production and productivity Measurement in Surveys. *Food Policy*. (online version) <https://doi.org/10.1016/j.foodpol.2021.102033>
- Kilic, T. and T. Sohnesen (2019). Same Questions but Different Answer: Experimental Evidence on Questionnaire design's Impact on Poverty Measured by Proxies. *Review of Income and Wealth* 65 (1).
- Kilic, T., Van den Broeck, G., Koolwal, G., & Moylan, H. (2020a). *Are You Being Asked? Impacts of Respondent Selection on Measuring Employment*. Tech. rep., The World Bank.
- 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
- Kish, L., & Collection, K. W. (1965). *Survey Sampling*. Wiley.
- Knippenberg, E., Jensen, N., & Constan, M. (2019). Quantifying household resilience with high frequency data : Temporal dynamics and methodological options. *World Development*, 121, 1–15. <https://doi.org/10.1016/j.worlddev.2019.04.010>
- Kosmowski, F., Chamberlin, J., Ayalew, H., Sida, T., Abay, K., & Craufurd, P. How accurate are yield estimates from crop cuts? Evidence from smallholder maize farms in Ethiopia, *Food Policy*, 102, 102122. <https://doi.org/10.1016/j.foodpol.2021.102122>.
- Kosmowski, F., & Worku, T. (2018). Evaluation of a miniaturized NIR spectrometer for cultivar identification: The case of barley, chickpea and sorghum in Ethiopia. *PLOS ONE*. doi:10.1371/journal.pone.0193620
- Kosmowski, F., Abebe, A., & Ozkan, D. (2020). Challenges and lessons for measuring soil metrics in household surveys. *Geoderma*. doi:10.1016/j.geoderma.2020.114500
- Kosmowski, F., Aragaw, A., Kilian, A., Ambel, A., Ilukor, J., YIGEZU, B. I., & Stevenson, J. (2019). Varietal identification in household surveys: results from three household-based methods against the benchmark of dna fingerprinting in southern Ethiopia. *Experimental Agriculture*, 55, 1-15. doi:10.1017/S0014479718000030
- Kretzschmar, T., Mbanjo, G., Magalit, G., Dwiyantri, M., Habib, M., Diaz, M., . . . Yamano, T. (2018). DNA fingerprinting at farm level maps rice biodiversity across Bangladesh and reveals regional varietal preferences. *Scientific Reports*, 8. doi:10.1038/s41598-018-33080-z
- Kreuter, F. (2013). *Improving Surveys with Paradata: Analytic Uses of Process Information*. John Wiley & Sons, Inc.
- Kristjanson, P., Waters-Bayer, A., Johnson, N., Tipilda, A., Njuki, J., Baltenweck, I., . . . Macmillan, S. (2014). Livestock and Women's Livelihoods. In A. Quisumbing, R. Meinzen-Dick, T. Raney, A. Croppenstedt, J. Behrman, & A. Peterman (Eds.), *Gender in Agriculture* (p. 20). doi:10.1007/978-94-017-8616-4_9.
- Krosnick, J. A. (1999). Survey research. *Annual review of psychology*, 50, 537–567.
- Laajaj, R., k. Macours, D. A. Pinzon Hernandez, O. Arias, S. Gosling, J. Potter, M. Rubio-Codina and R. Vakis. Challenges to Capture the Big Five Personality Traits in non-WEIRD Populations. *Science Advances*. Vol. 5. no. 7.

- Laajaj, R., & Macours, K. (2021). Measuring skills in developing countries. *Journal of Human resources*, 56 (4).
- LaFave, D., Peet, E., & Thomas, D. (2013). *Are rural markets complete? Prices, profits and recursion*. Tech. rep.
- Lamanna, C., Hachhethu, K., Chesterman, S., Singhal, G., Mwongela, B., Ngendo, M., Passeri, S., Farhikhtah, A., Kadiyala, S., Bauer, J. M., & Rosenstock, T. S. (2019). Strengths and limitations of computer assisted telephone interviews (CATI) for nutrition data collection in rural Kenya. *PLoS ONE*, 14(1), 1–20. <https://doi.org/10.1371/journal.pone.0210050>
- Lämmerhirt, D., Gray, J., Venturini, T., & Meunier, A. (2018). Advancing sustainability together? Citizen-generated Data and the sustainable development goals. *SSRN Electronic Journal*. doi:10.2139/ssrn.3320467
- Lavrakas, P. J. (2008). *Encyclopedia of survey research methods*. Sage Publishing. doi:10.4135/9781412963947
- Lavrakas, P. J. (2011). Is the Exclusion of Mobile Phones from Telephone Surveys a Problem: The U.S. Experience. *Presentation prepared for the Australian Mobile Phone Survey Workshop. Melbourne, Australia*.
- Leo, B., Morello, R., Mellon, J., Peixoto, T., & Davenport, S. T. (2015). Do Mobile Phone Surveys Work in Poor Countries? *Center for Global Development Working Paper No. 398*. <https://doi.org/10.2139/ssrn.2623097>
- Lesnoff, M., Lancelot, R., Moulin, C.-H., Messad, S., Juanès, X., & Sahut, C. (2014). *Calculation of Demographic Parameters in Tropical Livestock Herds*. doi:10.1007/978-94-017-9026-0
- Liao, C., Clark, P. E., DeGloria, S. D., & Barrett, C. B. (2017). Complexity in the spatial utilization of rangelands: Pastoral mobility in the Horn of Africa. *Applied Geography*, 86, 208-219. doi:<https://doi.org/10.1016/j.apgeog.2017.07.003>
- Liao, C., Clark, P., Shibia, M., & Degloria, S. (2018). Spatiotemporal dynamics of cattle behavior and resource selection patterns on East African rangelands: evidence from GPS-tracking. *International Journal of Geographical Information Science*, 32, 1-18. doi:10.1080/13658816.2018.1424856
- Lipper, L., McCarthy, N., Zilberman, D., Asfaw, S., & Branca, G. (2017). *Climate smart agriculture: building resilience to climate change*. Springer Nature.
- Little, P., Mcpeak, J., Barrett, C., & Kristjanson, P. (2008). Challenging Orthodoxies: Understanding Poverty in Pastoral Areas of East Africa. *Development and Change*, 39, 587-611. doi:10.1111/j.1467-7660.2008.00497.x
- Lobell, D., & Asseng, S. (2017). Comparing estimates of climate change impacts from process-based and statistical crop models. *Environmental Research Letters*, 12, 015001. doi:10.1088/1748-9326/aa518a
- Lobell, D., Deines, J., & Tommaso, S. (2020). Changes in the drought sensitivity of US maize yields. *Nature Food*, 1, 1-7. doi:10.1038/s43016-020-00165-w
- 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. doi:10.3390/rs12010100
- Lyberg, L., & Kasprzyk, D. (2004). Data Collection Methods and Measurement Error: An Overview. In P. P. Biemer, R. M. Groves, L. E. Lyberg, N. A. Mathiowetz, & S. Sudman (Eds.), *Measurement Errors in Surveys* (pp. 235-257). John Wiley & Sons, Ltd. doi:<https://doi.org/10.1002/9781118150382.ch13>

- MacDonald, J.M. (2016). *Structural Transformation in North America: What Does it Mean for Agricultural Statistics?* Seventh International Conference on Agricultural Statistics, Rome, Italy.
- Mahfoud, Z., Ghandour, L., Ghandour, B., Mokdad, A., & Sibai, A. (2014). Cell Phone and Face-to-face Interview Responses in Population-based Surveys: How Do They Compare? *Field Methods*, 27, 39-54. doi:10.1177/1525822X14540084
- Manski, C., & Molinari, F. (2008). Skip sequencing: A decision problem in questionnaire design. *The annals of applied statistics*, 2, 264-285. doi:10.1214/07-AOAS134
- Maredia, M., Reyes, B., Manu-Aduening, J., Dankyi, A., Hamazakaza, P., Muimui, K., . . . Raatz, B. (2016). Testing Alternative Methods of Varietal Identification Using DNA Fingerprinting: Results of Pilot Studies in Ghana and Zambia. *Testing Alternative Methods of Varietal Identification Using DNA Fingerprinting: Results of Pilot Studies in Ghana and Zambia*. doi:10.13140/RG.2.2.11573.27361
- Masuda, Y., Kelly, A., Robinson, B., Holland, M., Bedford, C., Childress, M., . . . Veit, P. (2020). How do practitioners characterize land tenure security? *Conservation Science and Practice*, 2. doi:10.1111/csp2.186
- Mathiowetz, N. A. (2000). The effect of length of recall on the quality of survey data. *Fourth Conference on Methodological Issues in Official Statistics, Stockholm, Sweden*.
- Maue, C. C., M. Burke and K.J. Emerick (2020). Productivity Dispersion and Persistence Among the World's Most Numerous Firms. *NBER Working paper No. 26924*
- McCarthy, N. (2011). Understanding agricultural households' adaptation to climate change and implications for mitigation: land management and investment options. *LSMS Guidebook*. Washington, DC: World Bank.
- McCarthy, N., Dutilly-Diane, C., Drabo, B., Kamara, A., & Vanderlinden, J.-P. (2004). Managing resources in erratic environments: An analysis of pastoralist systems in Ethiopia, Niger and Burkina Faso. *Research Report of the International Food Policy Research Institute*.
- McCullough, E. (2016). Labor productivity and employment gaps in Sub-Saharan Africa. *Food Policy*, 67. doi:10.1016/j.foodpol.2016.09.013
- Mercer, A. (2012) Using Paradata to Understand Effort and Attrition in a Panel Survey. *Section on Survey Research Methods – JSM*. 3822-3833.
- Meyer, B., Mok, W., & Sullivan, J. (2015). Household Surveys in Crisis. *Journal of Economic Perspectives*, 29, 199-226. doi:10.1257/jep.29.4.199
- Michelson, H., Fairbairn, A., Ellison, B., Maertens, A., & Manyong, V. (2021). Misperceived quality: Fertilizer in Tanzania. *Journal of Development Economics*, 148. doi:https://doi.org/10.1016/j.jdeveco.2020.102579
- Michler, O., Decker, R., & Stummer, C. (2019). To trust or not to trust smart consumer products: a literature review of trust-building factors. *Management Review Quarterly*, 70. doi:10.1007/s11301-019-00171-8
- Millán, T., Barham, T., Macours, K., Maluccio, J., & Stampini, M. (2019). Long-Term Impacts of Conditional Cash Transfers: Review of the Evidence. *The World Bank Research Observer*, 34, 119. doi:10.1093/wbro/lky005
- Minet, J., Curnel, Y., Gobin, A., Goffart, J. P., Mélard, F., Tychon, B., . . . Defourny, P. (2017). Crowdsourcing for agricultural applications: A review of uses and opportunities for a farmsourcing approach. *Computers and Electronics in Agriculture*, 142, Part A, 126-138. doi:10.1016/j.compag.2017.08.026

- Minten, B., Beyene, S., Legesse, E., & Kuma, T. (2015). Transforming Staple Food Value Chains in Africa: The Case of Teff in Ethiopia. *The Journal of Development Studies*, 52, 1-19. doi:10.1080/00220388.2015.1087509
- Minten, B., S. Tamru, E. Engida and T. Kuma (2016). Transforming Staple Food Value Chains in Africa: the Case of Teff in Ethiopia. *Journal of Development Studies* 52(5).
- Moore, J. C. (1988). Self/proxy response status and survey response quality. *Journal of Official Statistics*, 4, 155-172.
- Mundlak, Y. (2001). Chapter 1 Production and supply. *Handbook of Agricultural Economics*, 1, 3-85. doi:10.1016/S1574-0072(01)10004-6
- Muyanga, M., & Jayne, T. S. (2019). Revisiting the Farm Size-Productivity Relationship Based on a Relatively Wide Range of Farm Sizes: Evidence from Kenya. *American Journal of Agricultural Economics*, 101, 1140-1163.
- National Academies of Sciences, Engineering and Medicine. (2019). *Improving Data Collection and Measurement of Complex Farms*. Kling, C. and C. Mackie, eds. Washington, DC: The National Academies Press. <https://doi.org/10.17226/25260>
- Neter, J., & Waksberg, J. (1964). A Study of Response Errors in Expenditures Data from Household Interviews. *Journal of the American Statistical Association*, 59, 18-55.
- Nguyen, G., & Nguyen, T. T. (2020). Exposure to weather shocks: A comparison between self-reported record and extreme weather data. *Economic Analysis and Policy*, 65, 117-138. doi:10.1016/j.eap.2019.11.009
- Nicolas, G., Robinson, T. P., Wint, G. W., Conchedda, G., Cinardi, G., & Gilbert, M. (2016). Using random forest to improve the downscaling of global livestock census data. *PloS one*, 11.
- Nord, A., & Snapp, S. (2020). Documentation of farmer perceptions and site-specific properties to improve soil management on smallholder farms in Tanzania. *Land Degradation & Development*, 31, 2074-2086. doi:<https://doi.org/10.1002/ldr.3582>
- Norton, B.P., Hoel, J.B. & Michelson, H. (2020). The demand for (fake?) fertilizer: Using an experimental auction to examine the role of beliefs on agricultural input demand in Tanzania. 2020 Annual Meeting, July 26-28, Kansas City, Missouri 304444, Agricultural and Applied Economics Association.
- O'Sullivan, M., Rao, A., Banerjee, R., Gulati, K., & Vinez, M. (2014). Levelling the Field: Improving Opportunities for Women Farmers in Africa. *World Bank and One Campaign, Washington, DC*.
- Ochieng, D. O., & Baulch, B. (2020). *Report on a study to crowdsource farmgate prices for maize and soybeans in Malawi*. MaSSP reports, International Food Policy Research Institute (IFPRI).
- Olsen, R. (2005). The Problem of Respondent Attrition: Survey Methodology is Key. *Monthly Labor Review*, 128.
- Oseni, G., Durazo, J., & McGee, K. (2017). *The Use of Non-Standard Units for the Collection of Food Quantity : A Guidebook for Improving the Measurement of Food Consumption and Agricultural Production in Living Standards Surveys*. Washington, DC: World Bank.
- Outes-Leon, I., & Dercon, S. (2008). Survey attrition and attrition bias in Young Lives.
- Parkes, B., Higginbottom, T., Hufkens, K., Ceballos, F., Kramer, B., & Foster, T. (2019). Weather dataset choice introduces uncertainty to estimates of crop yield responses to climate variability and change. *Environmental Research Letters*, 14. doi:10.1088/1748-9326/ab5ebb

- Payne, S. L. (1980). *The Art of Asking Questions: Studies in Public Opinion*, 3 (Vol. 451). Princeton University Press.
- Pelletier, J., H. Ngoma, N.M. Mason and C.B. Barrett (2020). Does Smallholder Maize Intensification Reduce Deforestation? Evidence from Zambia. *Global Environmental Change*. 63.
- Pica-Ciamarra, U., Baker, D., Morgan, N., Zezza, A., Azzarri, C., Ly, C., . . . Sserugga, J. (2014). Investing in the livestock sector: Why good numbers matter, a sourcebook for decision makers on how to improve livestock data.
- Pischke, J.-S. (1995). Individual Income, Incomplete Information, and Aggregate Consumption. *Econometrica*, 63, 805–840.
- 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.
- Ponzini, G. et al. (2021). *Documenting the Uganda experience in the 50 by 2030 Initiative*. Unpublished manuscript. World Bank and FAO: Washington, DC and Rome. unpublished.
- Pope, R. D., & Just, R. E. (2003). Distinguishing Errors in Measurement from Errors in Optimization. *American Journal of Agricultural Economics*, 85, 348-358. doi:10.1111/1467-8276.00124
- Porciello, J., Ivanina, M., Islam, M., Einarson, S., & Hirsh, H. (2020). Accelerating evidence-informed decision-making for the Sustainable Development Goals using machine learning. *Nature Machine Intelligence*, 2, 559-565. doi:10.1038/s42256-020-00235-5
- Pratt, M., Sallis, J. F., Cain, K. L., Conway, T. L., Palacios-Lopez, A., Zezza, A., . . . Kilic, T. (2020). Physical activity and sedentary time in a rural adult population in Malawi compared with an age-matched US urban population. *BMJ Open Sport & Exercise Medicine*, 6. doi:10.1136/bmjsem-2020-000812
- Reardon, T., & Glewwe, P. (2000). "Agriculture" and "Module for Chapter 19" in M. Grosh and P. Glewwe (eds.) *"Designing household survey questionnaires for developing countries: lessons from 15 years of the living standards measurement study"*. World Bank.
- Reinermann, S., Asam, S., & Kuenzer, C. (2020). Remote Sensing of Grassland Production and Management—A Review. *Remote Sensing*, 12. doi:10.3390/rs12121949
- Řezník T, Pavelka T, Herman L, Lukas V, Širůček P, Leitgeb Š, & Leitner F. (2020.) Prediction of Yield Productivity Zones from Landsat 8 and Sentinel-2A/B and Their Evaluation Using Farm Machinery Measurements. *Remote Sensing*. 12(12):1917. <https://doi.org/10.3390/rs12121917>
- Ridolfo, H., D. Biagas, E. J. Abayomi and J. Rodhouse (2021). Behavior Coding of the 2018 Agricultural Labor Survey. *RDD Research Report No. RDD-21-01*, NASS, Washington, DC.
- Roberts, M. J., Schlenker, W., & Eyer, J. (2013). Agronomic Weather Measures in Econometric Models of Crop Yield with Implications for Climate Change. *American Journal of Agricultural Economics*, 95, 236-243. doi:<https://doi.org/10.1093/ajae/aas047>
- Robinson, T. P., Wint, G. W., Conchedda, G., Van Boeckel, T. P., Ercoli, V., Palamara, E., . . . Gilbert, M. (2014). Mapping the global distribution of livestock. *PloS one*, 9, e96084.
- Rodhouse, J., H. Ridolfo, E. Abayomi and D. Biagas (2019). Did the Respondent Really mean That? How the Behaviors of CATI Interviewers and Data Editors Impact Measurement and PProcessing Errors in Establishment Surveys. *2019 Workshop: Interviewers and*

- Their Effects from a Total Survey Error Perspective.* 35.
<http://digitalcommons.unl.edu/sociw/35>
- Rosenzweig, M. (2003). Payoffs from Panels in Low-Income Countries: Economic Development and Economic Mobility. *American Economic Review*, 93, 112-117.
doi:10.1257/000282803321946903
- Rosenzweig, M. R., & Udry, C. (2014). *Rainfall Forecasts, Weather and Wages over the Agricultural Production Cycle*. Working Paper, National Bureau of Economic Research.
doi:10.3386/w19808
- Roßmann J, Gummer T. Using Paradata to Predict and Correct for Panel Attrition. (2016.) *Social Science Computer Review*. 34(3):312-332. doi:10.1177/0894439315587258
- Rubin, D. B. (1987). *Multiple Imputation for Nonresponse in Surveys*. Wiley.
- Rubin, D. B. (1996). Multiple Imputation after 18+ Years. *Journal of the American Statistical Association*, 91, 473-489. doi:10.1080/01621459.1996.10476908
- Sagesaka, A., Palacios-Lopez, A., & Amankwah, A. (2020). *Measuring Work on Household Farms using Household Surveys. A Practical Guidebook on Designing Household Surveys for Effective Data Collection of Work on Household Farms*. Tech. rep., World Bank: Washington, DC.
- Salganik, M. J. (2019). *Bit by bit: Social research in the digital age*. Princeton University Press.
- Schennach, S. (2016). Recent Advances in the Measurement Error Literature. *Annual Review of Economics*, 8, 341-377. doi:10.1146/annurev-economics-080315-015058
- Schennach, S. M. (2004). Estimation of Nonlinear Models with Measurement Error. *Econometrica*, 72, 33-75.
- Schuman, H. and S. Presser (1981). *Questions and Answers in Attitude Surveys*. New York: Academic Press
- Schündeln, M. (2018). Multiple Visits and Data Quality in Household Surveys. *Oxford Bulletin of Economics and Statistics*, 80, 380-405. doi:<https://doi.org/10.1111/obes.12196>
- Schwarz, N. (1997). Questionnaire Design: the Rocky Road from Concept to Answer. In *Survey Measurement and Process Quality*, L. Lyberg et al. New York: John Wiley and Sons.
- Schwarz, N. and H. Hipplker (1991). Response alternatives: the impact of their choice and presentation order. in *Measurement Errors in Surveys*, P. Biemer and others, eds. New York: John Wiley and Sons.
- Scott, C. and B. Amenuvegbe (1991). Recall Loss and Recall Duration: and Experimental Study in Ghana. *Inter-Stat* 4(1)
- Sherlund, S., C. B. Barrett and A. Adesina (2002). Smallholder Technical Efficiency Controlling for Environmental Production Conditions. *Journal of Development Economics*. 69.
- Silberstein, A. R., & Scott, S. (2004). Expenditure Diary Surveys and Their Associated Errors. In P. P. Biemer, R. M. Groves, L. E. Lyberg, N. A. Mathiowetz, & S. Sudman (Eds.), *Measurement Errors in Surveys* (pp. 303-326). John Wiley & Sons, Ltd.
doi:<https://doi.org/10.1002/9781118150382.ch16>
- Singh, I., Squire, L., & Strauss, J. (1986). *Agricultural household models: Extensions, applications, and policy*. Washington, DC: World Bank.
- Sinha, P., Robson, A., Schneider, D., Kilic, T., Mugera, H. K., Ilukor, J., & Tindamanyire, J. M. (2020). The potential of in-situ hyperspectral remote sensing for differentiating 12 banana genotypes grown in Uganda. *ISPRS Journal of Photogrammetry and Remote Sensing*, 167, 85-103. doi:<https://doi.org/10.1016/j.isprsjprs.2020.06.023>
- Slavec, A., & Toninelli, D. (2015). An Overview of Mobile CATI Issues in Europe. In D.

- Toninelli, R. Pinter, & P. de Pedraza (Eds.), *Mobile Research Methods: Opportunities and Challenges of Mobile Research Methodologies* (pp. 41–62). Ubiquity Press.
- Song, L., Jiang, Q., Shi, Y.-E., Feng, X.-T., Li, Y., Su, F., & Liu, C. (2018). Feasibility Investigation of 3D Printing Technology for Geotechnical Physical Models: Study of Tunnels. *Rock Mechanics and Rock Engineering*, 51. doi:10.1007/s00603-018-1504-3
- Stajanko, D., Brus, M., & Hočevár, M. (2008). Estimation of bull live weight through thermographically measured body dimensions. *Computers and Electronics in Agriculture*, 61, 233-240. doi:10.1016/j.compag.2007.12.002
- Stewart, N., Chandler, J., & Paolacci, G. (2017). Crowdsourcing Samples in Cognitive Science. *Trends in Cognitive Sciences*, 21. doi:10.1016/j.tics.2017.06.007
- Strack F., L. Martin and N. Schwarz (1988). Priming and Communication: Social Determinants of Information Use in Judgement of Life Satisfaction. *European Journal of Social Psychology*. 18 (3).
- Sudman, S., & Bradburn, N. M. (1973). Effects of Time and Memory Factors on Response in Surveys. *Journal of the American Statistical Association*, 68, 805-815. doi:10.1080/01621459.1973.10481428
- Sudman, S., & Bradburn, N. M. (1974). *Response Effects in Surveys: Review and Synthesis*. Chicago: Aldine.
- Sudman, S., N. Bradburn and N. Schwarz (1996). *Thinking about Answers: the Application of Cognitive Processes to Survey Methodology*. San Francisco, CA: Jossey-Bass
- Swain, D., Friend, M., Bishop-Hurley, G. J., Handcock, R., & Wark, T. (2011). Tracking livestock using global positioning systems - are we still lost? *Anim. Prod. Sci.*, 51, 167-175. doi:10.1071/an10255
- Thomas, D., Frankenberg, E., & Smith, J. P. (2001). Lost but Not Forgotten: Attrition and Follow-up in the Indonesia Family Life Survey. *The Journal of Human Resources*, 36, 556–592.
- Thomas, D., Witoelar, F., Frankenberg, E., Sikoki, B., Strauss, J., Sumantri, C., & Suriastini, W. (2012). Cutting the costs of attrition: Results from the Indonesia Family Life Survey. *Journal of Development Economics*, 98, 108-123. doi:10.1016/j.jdeveco.2010.08
- Tourangeau, R., L.J. Rips and K. Rasinski (Eds.) (2000). *The Psychology of Survey Response*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511819322>
- Tourangeau, R., & Yan, T. (2007). Sensitive questions in surveys. *Psychological bulletin*, 133 5, 859-83.
- Turner, L., Udall, M., Larson, B., & Shearer, S. (2000). Monitoring cattle behavior and pasture use with GPS and GIS. *Canadian Journal of Animal Science - CAN J ANIM SCI*, 80, 405-413. doi:10.4141/A99-093
- Udry, C. (1996). Gender, Agricultural Production, and the Theory of the Household. *Journal of Political Economy*, 104, 1010-1046.
- UN Global Pulse. (2015). *Feasibility Study: Crowdsourcing High- Frequency Food Price Data in Rural Indonesia*.
- United Nations. (2019). *Guidelines for Producing Statistics on Asset Ownership from a Gender Perspective*. United Nations, New York.
- United Nations, World Bank. 2020. *Monitoring the State of Statistical Operations under the COVID-19 Pandemic*. Washington, DC: World Bank.

- Vaessen, M at al. (1987). Translation of Questionnaires into Local languages. In *The World Fertility Survey: an Assessment*. J. Cleland and C. Scott, eds. New York: Oxford University Press.
- Van De Giesen, N., Hut, R., & Selker, J. (2014). The Trans-African Hydro-Meteorological Observatory (TAHMO). *Wiley Interdisciplinary Reviews: Water*, 1. doi:10.1002/wat2.1034
- Vasques, G., Rodrigues, H., Coelho, M., Baca, J., Dart, R., Oliveira, R., . . . Ceddia, M. (2020). Field Proximal Soil Sensor Fusion for Improving High-Resolution Soil Property Maps. *Soil Systems*, 4. doi:https://doi.org/10.3390/soilsystems4030052
- Vijverberg, W. P., & Mead, D. C. (2000). "Household Enterprises", pp. 105-137 in Margaret Grosh and Paul Glewwe(eds.) *"Designing Household Survey Questionnaires for Developing Countries : Lessons from 15 Years of the Living Standards Measurement Study, Volume 3"*. The World Bank.
- Waldman, K., V. N., Attari, S., Sheffield, J., Estes, L., Caylor, K., & Evans, T. (2019). Cognitive biases about climate variability in smallholder farming systems in Zambia. *Weather, Climate, and Society*, 11(2), 369-383. doi:https://doi.org/10.1175/WCAS-D-18-0050.1
- Walker, T. S., & Ryan, J. G. (1990). *Village and household economics in India's semi-arid tropics*. Johns Hopkins University Press.
- Wansbeek, T., & Meijer, E. (2000). Measurement Error and Latent Variables in Econometrics. *Economics Letters*, 69.
- West, B., & Blom, A. (2017). Explaining Interviewer Effects: A Research Synthesis. *Journal of Survey Statistics and Methodology*, 5, 175-211. doi:10.1093/jssam/smw024
- Wiggins, A., Greg, N., Stevenson, R., & Crowston, K. (2011). *Mechanisms for Data Quality and Validation in Citizen Science*. 2011 IEEE Seventh International Conference on se-Science Workshops, Stockholm, Sweden . doi:10.1109/eScienceW.2011.27
- Winkielman, P. B. Knauper and N. Schwarz (1988). Looking back at anger: reference periods change the interpretation of emotional frequency questions. *Journal of Personality and Social Psychology*. 75(3)
- Witoelar, F. (2011). *Tracking in Longitudinal Household Surveys*. Tech. rep., Washington, DC: World Bank.
- Wollburg, P., Tiberti, M., & Zezza, A. (2020). Recall length and measurement error in agricultural surveys. *Food Policy*, 102003. doi:https://doi.org/10.1016/j.foodpol.2020.102003
- Wooldridge, J. M. (2002). Inverse probability weighted M-estimators for sample selection, attrition, and stratification. *Portuguese Economic Journal*, 1, 117–139.
- Working, H. (1925). The Statistical Determination of Demand Curves. 39(4). doi:10.2307/1883264
- 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, 1-16.
- Xue, L., Liu, G., Parfitt, J., Liu, X., van Herpen, E., Stenmarck, Å., . . . Cheng, S. (2017). Missing Food, Missing Data? A Critical Review of Global Food Losses and Food Waste Data. *Environmental Science & Technology*, 51. doi:10.1021/acs.est.7b00401
- Yigezu, Y. A., Alwang, J., Rahman, W., Mollah, M. B., El-Shater, T., Aw-Hassan, A., & Sarker, A. (2018). Is DNA fingerprinting the gold standard for estimation of adoption and impacts of improved lentil varieties? *Food Policy*, 83. doi:10.1016/j.foodpol.2018.11.004

- Yirga, C., & Alemu, D. (2016). Adoption of Crop Technologies among Smallholder Farmers in Ethiopia: Implications for Research and Development. *Ethiopian Journal of Agricultural Science, Ethiop. J. Agric. Sci.*, 1-16.
- Young Lives. (2020). *Listening to Young Lives at Work Covid-19 Phone Survey: First Call shows widening inequality*. Young Lives.
<https://www.younglives.org.uk/content/listening-young-lives-work-covid-19-phone-survey-first-call-shows-widening-inequality>
- Zabel, J. E. (1998). An Analysis of Attrition in the Panel Study of Income Dynamics and the Survey of Income and Program Participation with an Application to a Model of Labor Market Behavior. *Journal of Human Resources*, 33, 479-506.
- Zeug, H., Zeug, G., Bielski, C., Solano-Hermosilla, G., Mâ, R., & others. (2017). Innovative Food Price Collection in Developing Countries. Focus on Crowdsourcing in Africa. *JRC Working Papers*. doi:10.2788/12343
- Zezza, A., Federighi, G., Kalilou, A. A., & Hiernaux, P. (2016a). Milking the data: Measuring milk off-take in extensive livestock systems. Experimental evidence from Niger. *Food Policy*, 59, 174-186. doi:<https://doi.org/10.1016/j.foodpol.2016.01.005>
- Zezza, A., Pica-Ciamarra, U., Mugera, K. H., Mwisomba, T., & Okello, P. (2016b). Measuring the Role of Livestock in the Household Economy. *A Guidebook for Designing Household Survey Questionnaires*. 67pp.
- Zhang, P., Zhang, J., & Chen, M. (2017). Economic impacts of climate change on agriculture: The importance of additional climatic variables other than temperature and precipitation. *Journal of Environmental Economics and Management*, 83, 8-31.