Journal of Development Economics 158 (2022) 102927

Contents lists available at ScienceDirect

Journal of Development Economics

journal homepage: www.elsevier.com/locate/devec

Regular article

Privacy protection, measurement error, and the integration of remote sensing and socioeconomic survey data☆

Jeffrey D. Michler [a,*], Anna Josephson [a], Talip Kilic [b], Siobhan Murray [b]

[a] Department of Agricultural and Resource Economics, University of Arizona, United States of America
[b] Development Data Group, World Bank, United States of America

ARTICLE INFO

Dataset link: https://doi.org/10.5281/zenodo.6841500

JEL classification:
C38
C81
D83
O13
Q12

Keywords:
Spatial anonymization
Privacy protection
Remote sensing data
Measurement error
Sub-Saharan Africa

ABSTRACT

When publishing socioeconomic survey data, survey programs implement a variety of statistical methods designed to preserve privacy but which come at the cost of distorting the data. We explore the extent to which spatial anonymization methods to preserve privacy in the large-scale surveys supported by the World Bank Living Standards Measurement Study-Integrated Surveys on Agriculture (LSMS-ISA) introduce measurement error in econometric estimates when that survey data is integrated with remote sensing weather data. Guided by a pre-analysis plan, we produce 90 linked weather-household datasets that vary by the spatial anonymization method and the remote sensing weather product. By varying the data along with the econometric model we quantify the magnitude and significance of measurement error coming from the loss of accuracy that results from privacy protection measures. We find that spatial anonymization techniques currently in general use have, on average, limited to no impact on estimates of the relationship between weather and agricultural productivity. However, the degree to which spatial anonymization introduces mismeasurement is a function of which remote sensing weather product is used in the analysis. We conclude that care must be taken in choosing a remote sensing weather product when looking to integrate it with publicly available survey data.

1. Introduction

Public use datasets from large-scale household surveys play a central role in tracking progress towards national and international development goals and in formulating a wide array of development research. These surveys include those that are supported by the World Bank's Living Standards Measurement Study (LSMS), the USAID-funded Demographic and Health Surveys (DHS), and UNICEF's Multiple Indicator Cluster Surveys (MICS). In making these datasets public, survey programs must balance the demand for accurate data with the need for privacy protection. The more accurate the public data, the more privacy is lost (Dinur and Nissim, 2003).

To preserve privacy when publishing data, survey programs implement statistical disclosure limitation (SDL). SDL methods distort data, preserving privacy but reducing data accuracy and interoperability, both key requirements for data to generate value for development (Jolliffe et al., 2021). Interoperability relates to the ease with which different data sources can be linked through various means, including geographic coordinates or common geographic identifiers. In large-scale household surveys, the use of Global Positioning System (GPS)

☆ A pre-analysis plan for this research has been filed with Open Science Framework (OSF): https://osf.io/8hnz5/. We gratefully acknowledge funding from the World Bank's Living Standards Measurement Study (LSMS) and Knowledge for Change Program (KCP). We owe a particular debt to Andrew Dillon, who in several conversations helped us articulate the privacy protection issues in this data and encouraged us to pursue this topic in more depth. This paper has been shaped by conversations with Leah Bevis, Aine McCarthy, and Emilia Tjernström as well as seminar participants at the 44th BREAD Conference on Development Economics at Northwestern University, the IFAD 2022 conference in Rome, the Nordic Conference in Development Economics 2022 in Helsinki, the Midwest International Development Conference 2022 in Minneapolis, the Centre for the Study of African Economies Conference 2022 held virtually, PacDev 2022 held virtually, and the Methods and Measurement Conference 2021 held virtually. Earlier drafts of this paper were presented under the title "Estimating the Impact of Weather on Agriculture" at the AAEA annual meetings in 2017 in Chicago and in 2019 in Atlanta, the 31st triennial ICAE conference held virtually, and in seminar presentations at Arizona State University, the University of Minnesota, the World Bank, and Virginia Tech. We are especially grateful to Alison Conley, Emil Kee-Tui, and Brian McGreal for their diligent work as research assistants and to Oscar Barriga Cabanillas and Aleksandr Michuda for early help in developing the Stata `wxsum` package.

* Corresponding author.
E-mail address: jdmichler@arizona.edu (J.D. Michler).

https://doi.org/10.1016/j.jdeveco.2022.102927
Received 10 January 2022; Received in revised form 25 May 2022; Accepted 23 June 2022
Available online 1 July 2022

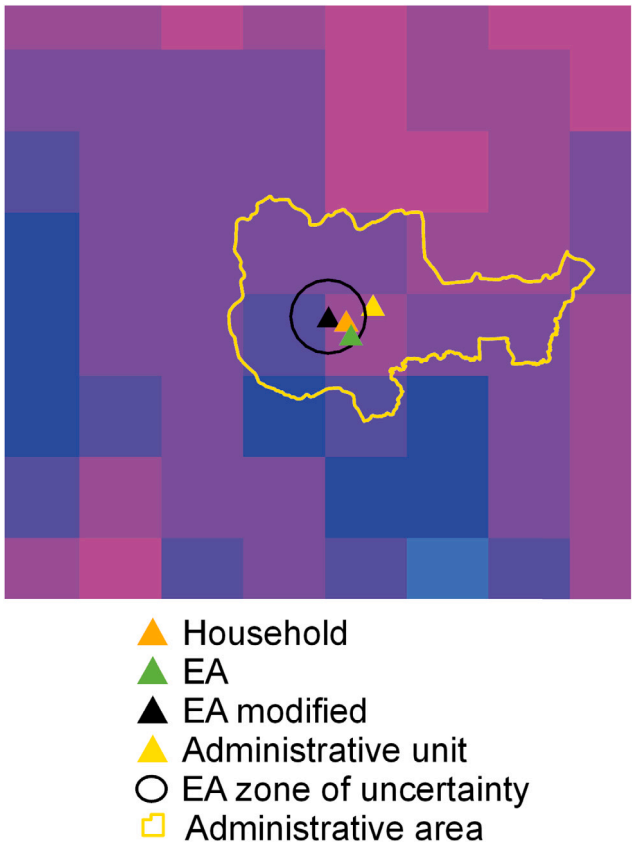

Fig. 1. Visualization of anonymization methods.
Note: The figure presents the different anonymization methods (see Table 4) and how the measurement of anonymization method would vary across a particular gridded remote sensing precipitation product.

technology to capture sampled enumeration area (EA), household, and agricultural plot locations has dramatically increased the interoperability of survey data by allowing the integration of survey data with remote sensing data (Burke et al., 2021). Although capturing precise GPS coordinates increases interoperability, and thus the relevance and cost-effectiveness, of household surveys, such data are confidential and must be "spatially anonymized" before public release. International survey programs have thus adopted SDL coordinate masking techniques such that public use datasets that include anonymized unit-record microdata are also inclusive of spatially anonymized GPS coordinates. While a range of coordinate masking techniques exist (see Fig. 1), the technique that is currently used by the DHS and LSMS randomly offsets precise EA coordinates by zero to two kilometers (km) in urban areas and two to five km in rural areas, with one percent of rural areas displaced up to ten km (Blankespoor et al., 2021).

This paper contributes to the nascent economics literature on privacy protection and statistical accuracy. We integrate nine remotely sensed geospatial weather datasets with georeferenced longitudinal household survey data that have been collected across six Sub-Saharan African countries under the World Bank LSMS-Integrated Surveys on Agriculture (LSMS-ISA) initiative. Prior to the integration process, we use the confidential household GPS coordinates to generate ten different spatial representations of the precise household locations. Linking the weather data to the household survey data using each of these ten spatial representations allows us to quantify the magnitude and significance of measurement error coming from the loss of accuracy that results from different SDL methods to protect privacy. We test this by modeling the relationship between weather and smallholder agricultural productivity, as measured through the LSMS-ISA-supported household surveys.[1] Our goal is to provide guidance to researchers looking to integrate geospatial data with socioeconomic survey data regarding the degree to which their results may be mismeasured due to privacy protection methods.

There are three headline findings from our research. First, we find that spatial anonymization techniques currently in general use, such as those currently employed by the LSMS and the DHS, have, on average, limited to no impact on estimates of agricultural productivity. At this time, the spatial resolution of publicly available remote sensing weather products are generally too coarse for any of the spatial anonymization methods to make a substantial difference in which pixel a household ends up in. The LSMS and DHS offset EA centerpoints by two to ten km, depending on if the EA is urban or rural. By contrast, the resolution of the publicly available remote sensing data we use is anywhere between 4.1×4.1 km to 69×55 km. Second, and not unexpectedly, the degree to which spatial anonymization introduces mismeasurement is a function of which remote sensing weather product is used in the analysis. Remote sensing products that merge gauge and satellite data, such as ARC2, CHIRPS, and TAMSAT, are seemingly of a high enough resolution to be sensitive to some spatial anonymization techniques.[2] Remote sensing products that rely on assimilation models, such as ERA5 and MERRA-2, or products that primarily rely on gauge data, such as CPC, are of a low enough resolution that commonly used spatial anonymization techniques have no discernible impact on estimates of agricultural productivity. Third, estimates of weather's impact on agricultural productivity are also a function of the remote sensing data source, regardless of the degree of/approach to spatial anonymization. The extent to which weather impacts agricultural productivity varies substantially both in sign, significance, and magnitude, across remote sensing weather data products for the same spatial anonymization technique. These results suggest the need for care when choosing a remote sensing data product to integrate with socioeconomic survey data, as results can vary depending on the spatial anonymization technique used to protect privacy and the choice of product.

As noted above, there is scope for the impact of spatial anonymization to vary in accordance with the measurement error in geospatial data sources that household survey data are linked to — in our case, remote sensing weather data. The goal of a remote sensing weather product is to document an objective fact: that is, the volume of precipitation or the temperature in a given location at a given time. Inaccuracies introduced by either the sensor (e.g., infrared, microwave, optical) or the algorithm used to convert sensor data into rainfall or temperature (e.g., reanalysis, interpolation) means remote sensing products may mismeasure the objective fact. Simply with respect to the "raw" weather data, there can be substantial variation in what a remote sensing product reports as the actual rainfall or temperature in a given location. Figs. 2 and 3 show this variation across six remote sensing precipitation products and three temperature products. One precipitation product reports rainfall of zero to five millimeters (mm) in the southeast corner of the grid cell while a different product reports 47–64 mm for the same location on the same day. Temperature also varies by remote sensing product, with one product reporting a maximum temperature of 23 °C while another reports the maximum temperature that day as 27 °C.

That variation exists not only in the spatial resolution of the remote sensing data but also in the precipitation and temperature reported by each product informed how we implemented our research design. First, we developed a pre-analysis plan and registered it at Open Science Framework (Michler et al., 2019b). While pre-analysis plans have

[1] In addition to being an area of research itself, agricultural production and productivity are often used to proxy for a variety of economic outcomes, including economic growth (Deschêne and Greenstone, 2007), intra-household bargaining power (Corno et al., 2020), and migration (Jayachandran, 2006).

[2] Section 3.1 includes a full description of each of these products.

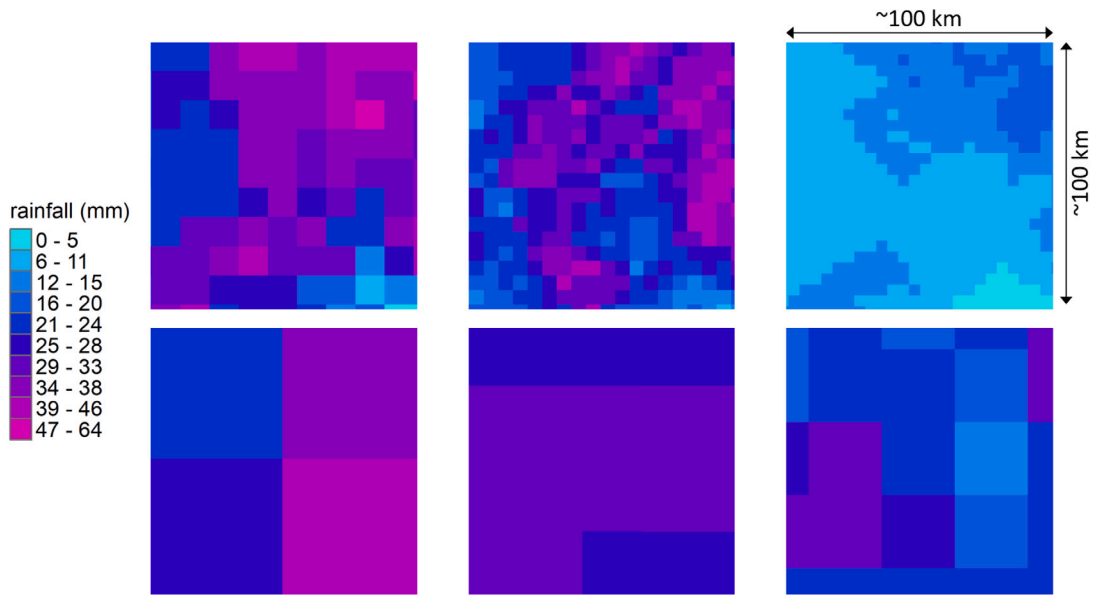

Fig. 2. Varying resolution of rainfall measurement.
Note: The figure captures rainfall as measured by all six precipitation products for the same 100 km × 100 km area on a single day (7 January 2010).

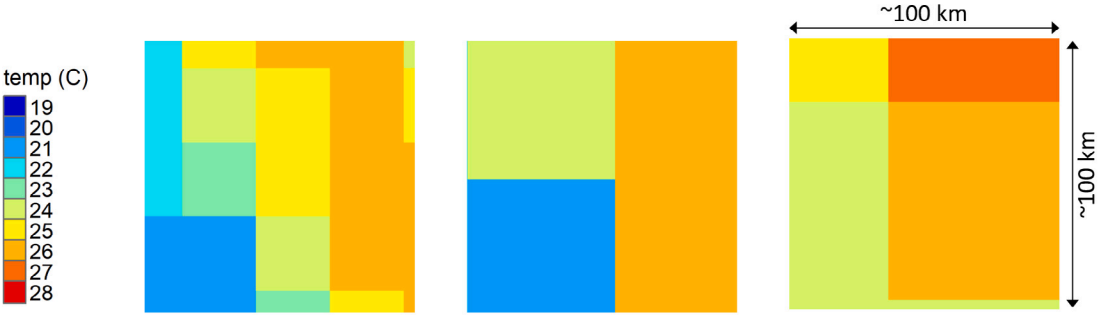

Fig. 3. Varying resolution of temperature measurement.
Note: The figure captures temperature as measured by all three temperature products for the same 100 km × 100 km area on a single day (7 January 2010).

become common in experimental economics, they are still relatively uncommon for binding researchers' hands when using observational data (Janzen and Michler, 2021). The use of a pre-analysis plan allowed us to pre-define the sources of data for inclusion in the study, what metrics would be tested using what functional forms, and how we would compare results across models in the absence of formal statistical tests. Second, we adopted a blinding strategy to help ensure objectivity in the implementation of the pre-analysis plan. As such, the authors were divided into two groups: the Data Generating Group and the Data Analysis Group. Authors Kilic and Murray were in the Data Generating Group and had full responsibility for extracting the remote sensing data and matching it to the household records in the household survey data to create a number of different paired weather-survey datasets.[3] In these datasets, the source of the weather data and the spatial anonymization method was anonymized prior to sharing with the Data Analysis Group. Authors Josephson and Michler made up the Data Analysis Group and had full responsibility for cleaning the agricultural productivity data, running the regressions, and conducting and writing the analysis. The pre-specified analysis was carried out on the blinded datasets and these results were posted to arXiv.org prior to unblinding (Michler et al., 2021b). The generation of datasets in this manner preserves the objectivity of any findings regarding differences in outcomes between different spatial anonymization techniques and different remote sensing products.

Against this background, this paper provides, to our knowledge, the first empirical evidence on the extent to which spatial anonymization of public use survey datasets affects econometric analysis when those datasets are linked to remote sensing data. We also provide evidence on how the significance and magnitude of the effect of spatial anonymization varies in accordance with the remote sensing data source. In our case, the unique access of the Data Generating Group to the confidential household GPS coordinates in the LSMS-ISA's nationally-representative, panel datasets allows us to execute the comparative assessment and isolate the role of spatial anonymization in the subsequent econometric analyses.

The issues surrounding privacy-preserving data analysis are well-known in computer science but have come to the widespread attention of economists only since the announcement by the US Census Bureau to implement differential privacy (DP) for the 2020 Census of Population (Abowd and Schmutte, 2019). The issue of accuracy in privacy-preserving data remains largely unexplored in the development economics literature, despite the proliferation of research on accuracy and measurement error in household survey data (Carletto et al., 2017; Abay et al., 2019; Kosmowski et al., 2019; Gollin and Udry, 2021; Kilic et al., 2021). To date, there is limited evidence on how the use of

[3] For example, in one dataset the remote sensing weather data product may be matched with the exact household coordinates, while in another dataset the remote sensing weather data may be matched with low-level Administrative area.

spatially anonymized public use datasets may impact the findings of research efforts that are centered on the integration of georeferenced socioeconomic survey data with satellite imagery and/or processed geospatial data. This is despite the rapid expansion in publicly available high-resolution satellite imagery, which has been used in combination with household survey data for small area estimation of poverty, wealth, health, nutrition, and agricultural outcomes in low-income contexts (Azzari et al., 2021; Burke and Lobell, 2017; Graetz et al., 2018; Osgood-Zimmerman et al., 2018; Dwyer-Lindgren et al., 2019; Yeh et al., 2020).

Relatedly, a large body of economic research has relied on remotely-sensed weather data for identification of causal effects (Dell et al., 2014; Donaldson and Storeygard, 2016). This includes important contributions that rely on the availability of georeferenced household survey data and that relate to human capital formation (Maccini and Yang, 2009; Shah and Steinberg, 2017; Garg et al., 2020), labor markets (Jayachandran, 2006; Chen et al., 2017; Kaur, 2019; Morten, 2019), conflict and institutions (Brückner and Ciccone, 2011; Sarsons, 2015; König et al., 2017), agricultural production and economic growth (Miguel et al., 2004; Deschêne and Greenstone, 2007; Barrios et al., 2010; Dell et al., 2012; Yeh et al., 2020), intra-household bargaining power (Corno et al., 2020), technology adoption (Suri, 2011; Taraz, 2018; Jagnani et al., 2021; Aragón et al., 2021; Tesfaye et al., 2021), and extreme weather impacts (Wineman et al., 2017; Michler et al., 2019a; McCarthy et al., 2021). Our findings suggest that economists should exercise caution when seeking to combine remote sensing data with public use socioeconomic survey data.

The remainder of the paper is organized as follows: in Section 2 we discuss the issue of privacy loss, different methods for privacy protection, and their implications for economic analysis. We also discuss the current coordinate masking techniques used by the DHS and the LSMS to ensure spatial anonymity in their published datasets. Section 3 details the sources and characteristics of the weather data and the household data used in this analysis. We provide details on how data was integrated, including specifics on how the blinded data was combined. The section concludes by presenting some descriptive evidence of mismeasurement in the remotely sensed weather data. Section 4 gives details of the pre-analysis plan, specifically our estimation strategy and approach to inference. Section 5 discusses results while Section 6 concludes with a set of recommended best practices for researchers looking to integrate remote sensing data with socioeconomic survey data.

2. Privacy protection in socioeconomic data

Socioeconomic data, including personal data and household survey data, are collected with the understanding that the identity of individual respondents will be protected when the data are disseminated or used in research. This is the case with the large, public use datasets most commonly used in development economics, including those made available by the LSMS, the DHS, and the MICS. Statistical disclosure limitation (SDL) methods such as noise infusion, aggregation, record swapping, or suppression may be employed to reduce the uniqueness of any single record in the sample and maintain confidentiality. In the spatial dimension, SDL is often achieved through coordinate masking and noise infusion on derived spatial variables. SDL inherently distorts the data, which can lead to bias in statistical analysis (Abowd et al., 2019). Because data providers do not publish SDL critical parameters, so as to reduce the potential for database reconstruction, it is not possible to determine the magnitude or direction of the bias (Abowd and Schmutte, 2015).

Regardless of the SDL methods employed to protect privacy, the Database Reconstruction Theorem demonstrates that publishing too many statistics too accurately from a confidential database exposes the entire database with near certainty (Dinur and Nissim, 2003). Additionally, the expanding availability of personal data that can be linked to survey data, as well as the wide availability of software and computational resources for mining these data, means that data de-identified via traditional SDL are vulnerable to re-identification via record linkage. In recent years, companies like Apple, Facebook, and Google have used differential privacy (DP) techniques in preserving privacy of user data (Wood et al., 2018). This is also the method adopted by the US Census Bureau in preparing the 2020 Census data for release (Abowd et al., 2019). DP techniques allow for the precise measurement of disclosure risk, thereby avoiding excessive data manipulation, while meeting anonymization objectives (Dwork et al., 2006). The use of DP, or any privacy protecting statistical technique, raises important questions about social choice, privacy protection, data accuracy, and the transparency and reproducibility of research. This is a debate which economists are just now beginning to enter.[4]

As of 2022, DP has only just begun to be adopted by the statistical agencies and the managers of the databases most commonly used by economists. This includes the US Census Bureau, which adopted DP for the 2020 census. Privacy in the Opportunity Atlas, which is published at the Census tract level, is also protected by methods that build on DP (Chetty and Friedman, 2019). However, to date, public use household survey datasets in development economics still rely on SDL to protect participant privacy. While DP may hold promise for future household survey data dissemination, in this analysis we make use of existing LSMS-ISA public datasets which rely on SDL to anonymize location data. In the remainder of this section, we detail the SDL methods currently used in the LSMS-ISA data in addition to the various methods we test in our analysis.

2.1. Geomasking in the LSMS and DHS

Spatial anonymization has dual objectives: (1) to provide a geographic reference that enables users to integrate information from spatial datasets into a household survey and, at the same time, (2) to preserve confidentiality of place, preventing re-identification of the location of survey respondents. Geomasking, or coordinate perturbation, serves to conceal the actual location and, when mask parameters are revealed, also enables users to incorporate uncertainty into spatial variables derived using the anonymized locations. The geomasking technique applied to LSMS-ISA public microdata is a type of SDL developed by the DHS Program and has been used in the dissemination of survey datasets since the early 2010s (Blankespoor et al., 2021).

Specifically, the coordinate modification strategy relies on noise infusion through random offset or perturbation of EA centerpoint coordinates (the average of sample household GPS locations by EA) within a specified range determined by an urban/rural classification. For urban areas, a range of zero to two km is used to offset the true EA centerpoint. For rural areas, where communities are further dispersed and risk of disclosure could be greater, a range of zero to five km is used to offset the true EA centerpoint. An additional zero to ten km offset is used for a small percentage (ranging from one to ten percent) of rural areas, effectively increases the known range for all rural points to ten km while introducing only a small amount of additional noise. The result is a set of coordinates, representative at the EA level, that fall within limits of accuracy known to the data user.[5]

With the geomasking method described, there is no guarantee that specific anonymization objectives are achieved. Further, this geomasking method does not take into account location-specific characteristics, other than official rural/urban classification. Adaptive approaches, where displacement is a function of site characteristics or the offset range is defined by a target population count, have been explored

[4] See the symposium at the 2019 AEA Annual Meeting (Abowd et al., 2019; Abraham, 2019; Chetty and Friedman, 2019; Ruggles et al., 2019).

[5] The modification strategy is adjusted to ensure households remain within the administrative district, i.e., the smallest political unit in the data.

by both the LSMS and DHS. An adaptive approach has the potential to avoid instances of excessive displacement in densely populated areas, as well as inadequate protection in sparsely populated areas. However, uncertainty in gridded population data inputs at large scale remains a barrier to implementation of the adaptive approach in many settings (Blankespoor et al., 2021). As a result, the strata-based method remains the primary spatial anonymization for dissemination of the LSMS datasets at this time.

2.2. Spatial feature representation

Most household survey datasets include location variables (e.g., region, district, or other place names), that define a base level of spatial disclosure risk. Any additional spatial information, including anonymized coordinates, allows for refinement of the anonymizing region, or area within which the survey respondent is known to reside. The trade-off for this increased exposure risk is an expected gain in the accuracy of derived spatial variables, such as precipitation or temperature. As the unit of analysis in many analyses – this one included – is the household, variables derived using exact household coordinates are assumed to contain the least amount of noise but produce the greatest risk of re-identification.

We conduct a comparative assessment of six spatial representations of household location that provide varying degrees of accuracy and spatial anonymity:

1. **Household**: the true household point locations as captured by enumerators using GPS devices.
2. **Enumerator Area (EA)**: the true centerpoint of an EA, where centerpoint is the average of sampled household locations within an EA.
3. **EA modified**: the EA centerpoint, but modified or offset using the LSMS and DHS geomasking technique described in the previous subsection.
4. **Administrative unit**: the geographic centerpoint of the administrative unit associated with lowest-level locality variable in the public microdata.
5. **EA zone of uncertainty**: the area (polygon) around a given EA that corresponds to the maximum possible offset for that EA. For urban EAs, this is a two km diameter circle around the true EA centerpoint. For rural EAs, it is a ten km diameter circle around the true EA centerpoint.
6. **Administrative area**: the geographic area (polygon) that is mapped by the political boundaries of the administrative unit.

Table 1 summarizes these spatial features and describes them in terms of the average displacement distance and a qualitative assessment of the impact on spatial disclosure risk associated with the dissemination of the spatial representation of household location.

The average point displacement, which could be viewed as representing potential mismeasurement in the derived variables, varies somewhat by country and strata, depending on factors such as the areal extent of EAs and administrative units. However, the direction and magnitude of difference between feature types is common across all surveys in the analysis. While the effect of displacement distance may be generally progressive for landscape-level phenomena like weather from a medium resolution dataset, this impact is scale-dependent. One could expect that hyperlocal characteristics, like field-level vegetation indices, from a high resolution image, would be rendered unusable by insertion of almost any noise.

2.3. Extraction method

The spatial features discussed above are a mix of point and polygon, or area, representations (see Fig. 1). In this analysis we make use of multiple gridded, or raster, weather data sources produced at different spatial resolutions (see Figs. 2 and 3). The method by which raster values are linked to different spatial features can compensate to some degree for differences in feature size and grid resolution. For example, the EA zone of uncertainty or Administrative area may be smaller than a single grid cell or cover multiple cells. A point feature may lie on the boundary of two grid cells or be located near a cell center. Extraction method refers to the way underlying grid cell values are processed.

We evaluate three commonly employed techniques for merging values from raster data to household roster records using the six spatial representations of household location. For the four point locations we extract weather time series data using both simple and bilinear methods, resulting in eight outputs. The simple method extracts raster cell values by spatial intersection alone, not accounting for the point location within cell boundaries. The bilinear method computes the distance weighted average of values at the four nearest cell centers. It is important to note that the bilinear method is generally preferred for integration of continuous data like precipitation and temperature. However, as we are aiming to assess the added value of the more complex calculations in this context, both bilinear and simple are considered in our analysis. For the two polygon locations we extract values using a zonal mean, or average of all cells overlapped by the polygon. The use of polygon features can account for uncertainty in location, as with the EA zone of uncertainty or Administrative area. Zonal means will also smooth the results, reducing the effect of extreme cell values.[6]

Altogether, the combination of spatial feature representations and extraction methods gives us ten spatial representations of household location. In the following analysis we treat the true household coordinated extracted using the bilinear method (Household bilinear) as the "true" or exact household location and test the other nine methods against Household bilinear.[7] To reiterate, the LSMS-ISA public datasets include EA modified centerpoint coordinates.[8]

3. Data

To understand the privacy/accuracy trade-off in anonymizing spatial data, we combine publicly available satellite-based weather data products with publicly available unit-record survey data that have been generated as part of the World Bank LSMS-ISA initiative and that are made available through the World Bank Microdata Library. In this section, we first describe the weather data and household data. We then discuss the blinding of the research team and the data integration process. We conclude with a discussion of some descriptive statistics for the combined weather-household datasets.

3.1. Remote sensing weather data

We use a number of public domain sources of weather datasets representing different modeling types, input sources, and spatial resolutions. Although there are many possible weather products to consider, we sought to include the remote sensing data products most commonly used by economists. To ensure consistency and enable the production of common metrics across the analysis, we imposed two inclusion criteria. The source had to have (1) high temporal resolution, i.e., daily, and (2) a minimum 30-year length of record, from 1987 to, at least, 2017. Unfortunately, this criteria meant that some data sources frequently used by economists, including the various versions of the monthly

[6] Figure A1 in Online Appendix A provides a visual representation of these three different methods.

[7] In subsequent figures we visually highlight the results from Household bilinear using boldface text, red reference lines, or orange reference markers.

[8] Geovariables disseminated with the microdata are currently generated using the EA modified centerpoint location and bilinear extraction, unless the underlying spatial dataset is categorical, in which case the simple extraction method is used.

Table 1
Spatial anonymization method.

	Spatial feature	Extraction method	Anonymization approach	Displacement (km)	Spatial disclosure risk
Household	Point	Simple, bilinear	None	0.0	Enables household location identification
EA	Centerpoint	Simple, bilinear	Aggregation	0.5	High risk of community identification
EA modified	Centerpoint	Simple, bilinear	Aggregation + perturbation	2.0–10	Moderate risk of community identification
Administrative unit	Centerpoint	Simple, bilinear	Large area aggregation	16.8	No increase in risk if administrative unit is identified in microdata
EA zone of uncertainty	Polygon	Area mean	Aggregation + perturbation	N/A	Moderate risk of community identification
Administrative area	Polygon	Area mean	Large area aggregation	N/A	No increase in risk if administrative unit is identified in microdata

Note: The table summarizes the various spatial features and anonymization methods tested in the analysis. Households are represented by their point location recorded via GPS. EA, EA modified, and Administrative unit are represented by the centerpoint of the object/area. EA zone of uncertainty is the polygon enclosing the region around the centerpoint in which the centerpoint could be located (0–2 km for urban EAs, 0–10 km for rural EAs). Administrative area is the polygon that maps the political boundaries of the administrative unit. Points and centerpoints can be mapped onto gridded data in one of two ways: simple or bilinear. The simple method extracts the cell value in which a point falls. The bilinear method calculates the distance weighted average of values at the four nearest cell centers. For polygons, the average is taken of the cell values that fall within the polygon. Displacement is calculated as mean displacement distance from household location for all households with GPS in baseline wave.

Terrestrial Air Temperature and Precipitation from the Center for Climatic Research at the University of Delaware was excluded. Table 2 describes each data sources, including the length of record, spatial and temporal resolution, and the type of data recorded. See online Appendix A for more details on each remote sensing product and guidance for economists on merging these data with survey data.

The remote sensing weather data that we use can be categorized by its method of generating precipitation and temperature values. The first type of product we use merges gauge data, which provide site-level observations, with data from meteorological satellites, which provide valuable indirect information at full coverage. Remote sensing products of this type include the African Rainfall Climatology version 2 (ARC2), the Tropical Applications of Meteorology using SATellite data and ground-based observations (TAMSAT), and the Climate Hazards group InfraRed Precipitation with Station data (CHIRPS) (Novella and Thiaw, 2013; Tarnavsky et al., 2014; Funk et al., 2015).

The second type of product uses assimilation models to combine a large number of observations from different sources (e.g., satellites, weather stations, ships, aircraft) to produce a model of the global climate system or a particular atmospheric phenomenon. Outputs are inferred or predicted based on the system state and understanding of interactions between model variables. We use two reanalysis datasets for both rainfall and temperature: the European Centre for Medium-Range Weather Forecasts ERA5 and the NASA Modern-Era Retrospective analysis for Research and Applications (MERRA-2) (Hennermann and Berrisford, 2020; Bosilovich et al., 2016).

Last, we consider a data product produced primarily from gauge data, using only spatial interpolation techniques to produce a continuous surface from observed measurements. The NOAA Climate Prediction Center (CPC) Unified Gauge-Based Analysis of Daily Precipitation and Temperature datasets were created using all information sources available at CPC and undergoes extensive pre-processing and cleaning, including comparison with contemporaneous data from satellite and other sources (Chen et al., 2008).

3.2. Household survey data

The World Bank Living Standards Measurement Study - Integrated Surveys on Agriculture (LSMS-ISA) is a household survey program that provides financial and technical assistance to national statistical offices in Sub-Saharan Africa for the design and implementation of national, multi-topic longitudinal household surveys with a focus on agriculture. As detailed below, our analysis leverages data from several rounds of panel household surveys conducted over the last decade in Ethiopia, Malawi, Niger, Nigeria, Tanzania, and Uganda. Table 3 provides a summary of the countries, years, and observations used in the analysis. Online Appendix B provides greater details on each country's sampling frame and data collection process.

In Ethiopia, we use the data from the 2011/12, 2013/14 and 2015/16 rounds of the Ethiopia Socioeconomic Survey, which has been conducted by the Central Statistical Agency of Ethiopia (Central Statistics Agency of Ethiopia (CSA), 2014, 2015, 2017). The Wave 1 data is representative at the regional level for the most populous regions in the country while Wave 2 and 3 expanded to include 1500 households in urban areas. After data cleaning to remove urban and non-agricultural rural households, we are left with 7272 household observations across three survey waves.

In Malawi, the LSMS-ISA data includes two separate surveys: the cross-sectional Integrated Household Survey, and the longitudinal Integrated Household Panel Survey (National Statistical Office (NSO), 2012, 2015, 2017). This analysis relies on the data from the IHPS, which is representative at the national-, urban/rural-, and regional-level. Data comes from 2010/11, 2013, and 2016/17. After data cleaning to remove tracked and non-agricultural households, we are left with 3250 household observations across three survey waves.

In Niger, we use two waves, the first from 2011 and the second from 2014 (Survey and Census Division, National Institute of Statistics, Niger (NIS), 2014, 2016). The sample is representative at the national and urban/rural-level. Data cleaning and removal of non-agricultural households gives us 3913 household observations across two survey waves.

In Nigeria, we use the data from the 2010/11, 2012/13, and 2015/16 rounds of the General Household Survey - Panel, which is representative at the national and urban/rural-level (National Bureau of Statistics (NBS), 2012, 2014, 2019). Data cleaning and removal of non-agricultural households yields 8,384 household observations across three survey waves.

In Tanzania, the data come from the 2008/09, 2010/11, and 2012/13 rounds of the Tanzania National Panel Survey (TNBS, 2011, 2012, 2015). The sample is representative for the nation, and provides estimates of key socioeconomic variables for mainland rural areas, Dar es Salaam, other mainland urban areas, and Zanzibar. Focusing on rural, crop producing households that do not move, we have 5669 household observations across three survey waves.

In Uganda, we use the data from the 2009/10, 2010/11, and 2011/12 rounds of the Uganda National Panel Survey (UBOS, 2019, 2014a,b). As with the other LSMS-ISA data, the Uganda sample was designed to be representative at the national-, urban/rural- and regional-level. We include 5250 household observations after cleaning and removing non-agricultural households.

For the analysis, we combine data from the six countries and all waves to generate a single cross-country panel dataset which includes 33,738 household observations. In estimation, we include two measures of agricultural productivity: yield (kg/ha) of the primary cereal crop and the value (2010 USD/ha) of all seasonal crop production on the farm.[9]

[9] In Ethiopia, Malawi, Nigeria, Tanzania, and Uganda the primary cereal crop is maize. In Niger the primary crop is millet. Millet is more drought tolerant than maize, so *a priori* we would expect rainfall in Niger to have less of an impact relative to the maize-focused countries.

Table 2
Sources of weather data.

Dataset	Length of record	Resolution (°)	≈Grid size (km)	Time step	Data	Units
Precipitation						
–Africa Rainfall Climatology version 2 (ARC2)	1983-current	0.1	11 × 11	Daily	Total precip	mm
–Climate Hazards group InfraRed Precipitation with Station data (CHIRPS)	1981-current	0.05	5.5 × 5.5	Daily	Total precip	mm
–CPC Global Unified Gauge-Based Analysis of Daily Precipitation	1979-current	0.5	55 × 55	Daily	Total precip	mm
–European Centre for Medium-Range Weather Forecasts (ECMWF) ERA5	1979-current	0.28	31 × 31	Hourly	Total precip	m
–Modern-Era Retrospective analysis for Research and Applications, version 2 (MERRA-2) Surface Flux Diagnostics	1980-current	0.625 × 0.5	69 × 55	Hourly	Rain rate	kg m^2 s^1
–Tropical Applications of Meteorology using SATellite data and ground-based observations (TAMSAT)	1983-current	0.0375	4.1 × 4.1	Daily	Total precip	mm
Temperature						
–CPC Global Unified Gauge-Based Analysis of Daily Temperature	1979-current	0.5	55 × 55	Daily	Min, max temp	C
–European Centre for Medium-Range Weather Forecasts (ECMWF) ERA5	1979-current	0.28	31 × 31	Hourly	Mean temp	K
–Modern-Era Retrospective analysis for Research and Applications, version 2 (MERRA-2) statD	1980-current	0.625 × 0.5	69 × 55	Daily	Mean temp	K

Note: The table summarizes the remote sensing sources and related details for precipitation and temperature data.

Table 3
Sources of household data.

Country	Survey name	Years	Original *n*	Final *n*
Ethiopia	Ethiopia Socioeconomic Survey (ERSS)	2011/2012	3,969	1,689
		2013/2014	5,262	2,865
		2015/2016	4,954	2,718
Malawi	Integrated Household Panel Survey (IHPS)	2010/2011	3,246	1,241
		2013	4,000	968
		2016/2017	2,508	1,041
Niger	Enquête Nationale sur les Conditions de Vie des Ménages et l'Agriculture (ECVMA)	2011	3,968	2,223
		2014	3,617	1,690
Nigeria	General Household Survey (GHS)	2010/2011	5,000	2,833
		2012/2013	4,802	2,768
		2015/2016	4,613	2,783
Tanzania	Tanzania National Panel Survey (TZNPS)	2008/2009	3,280	1,907
		2010/2011	3,924	1,914
		2012/2013	3,924	1,848
Uganda	Uganda National Panel Survey (UNPS)	2009/2010	2,975	1,704
		2010/2011	2,716	1,741
		2011/2012	2,850	1,805
Total	6 countries	17 waves	65,608	33,738

Note: The table summarizes the household data details for each country, per LSMS Basic Information Documents.

3.3. Data integration

Methods of data integration are often overlooked in the process of merging spatial data, in particular weather data, with household surveys. Publicly available datasets obfuscate the exact GPS coordinates of unit-records to ensure privacy. If underlying datasets are fairly smooth and areas of interest are small relative to the resolution of spatial data, then the effect of integration method could be negligible. However, this is not known and so our analysis sheds light on this privacy/accuracy trade-off.

As defined in our pre-analysis plan, the authors divided themselves into two groups to blind the Data Analysis Group from the identity of the spatial anonymization technique as well as the source of the remote sensing data (Michler et al., 2019b). The entire team participated in the development and registration of the pre-analysis plan, which included defining the remote sensing products to be used and the anonymization methods to be employed. At that point, the Data Generating Group accessed the publicly available remote sensing data for use in the study. They also used the privately available household coordinate data to generate the ten different sets of anonymization methods to be assessed. The actual GPS household location is not part of the publicly available LSMS-ISA data and is known only to a limited number of individuals at the World Bank.[10]

After pre-processing, the Data Generating Group extracted the relevant remote sensing data for the LSMS-ISA households based on the ten spatial anonymization methods for all remote sensing sources. This generated time series datasets of daily precipitation or temperature from January 1, 1983 until December 31, 2017. And so, for each of the 17 LSMS-ISA country-wave household datasets, this generated 90 remote sensing weather datasets (six precipitation sources + three temperature sources × ten anonymization methods). The time series weather datasets include daily observations and the unique household identifiers made part of the publicly available LSMS-ISA data. datasets were named and labeled `x0, . . . , x9` for each anonymization

[10] Note that we rely on household coordinates to test anonymization and not plot-level coordinates. In the LSMS-ISA, the average distance between household and plot is 1.3 km, which is much smaller than the highest resolution data set. So, matching the multiple plots a household operates would greatly increase the computational burden without adding any new information to the analysis, as the average plot would be in the same grid cell as the household.

Table 4
Data scope.

Countries (6)	Ethiopia, Malawi, Niger, Nigeria, Tanzania, Uganda
Weather Products (9)	Precipitation ARC2, CHIRPS, CPC, ERA5, MERRA-2, TAMSAT Temperature CPC, ERA5, MERRA-2
Anonymization methods (10)	Point (simple) Household, EA center, EA center modified, Administrative center Point (bilinear) Household, EA center, EA center modified, Administrative center Polygon (area mean) EA zone of uncertainty, Administrative area
Weather metrics (22)	14 rainfall 8 temperature
Dependent variables (2)	Value, quantity
Specifications (4)	Linear without household & year FEs, with household & year FEs Quadratic without household & year FEs, with household & year FEs

Note: The table summarizes the scope of the data across country, weather product, anonymization method, weather metric, dependent variable, and econometric specification.

method, `rf1, . . . , rf6` for each precipitation data source, and `tp1, . . . , tp3` for each temperature data source. These 1530 blinded datasets were then shared, via a secure server, with the Data Analysis Group.

The Data Analysis Group then processed each of the time series weather datasets using a user-written Stata package `wxsum`, which is available through Github. This package processes daily precipitation or temperature data and outputs up to 22 different weather metrics. See Table A.1 in online Appendix A for a complete list of weather metrics used in the analysis. These weather metrics from each of the 1530 weather datasets were then merged to the relevant country-wave LSMS-ISA dataset using the unique household identifier (90 weather datasets per country-wave dataset). All country-wave datasets containing the productivity data and the weather metrics from each remote sensing source and extraction method were then appended to create a single panel dataset covering all countries, waves, remote sensing sources, and anonymization methods. Table 4 summarizes the scope of the resulting data.

Following (Duflo et al., 2020), we have produced a "populated pre-analysis plan" that completely reproduces the results of all pre-specified analysis. After the Data Analysis Group conducted all of the analysis on the blinded dataset, they posted the populated pre-analysis plan to arXiv.org on 19 August 2021. That version of the populated pre-analysis plan (arXiv:2012.11768v2) refers to all results based on their randomly assigned identifier (`x0, . . . , x9`; `rf1, . . . , rf6`; and `tp1, . . . , tp3`). On 23 August 2021, the Data Generating Group shared the key so that the Data Analysis Group could de-anonymize the data. The populated pre-analysis plan was then updated to replace the randomly assigned identifiers with the actual anonymization methods and names of remote sensing sources (arXiv:2012.11768v3).[11] The current research paper presents the subset of the pre-specified results that focused on the issue of spatial anonymization.

3.4. Descriptive statistics

Our pre-analysis plan specifies that we will examine 22 different ways to measure precipitation and temperature in order to evaluate certain weather metrics are more or less accurate to spatial anonymization methods used to ensure participant privacy. A complete list of these variables with their exact definitions are in Table A.1 in online Appendix A. For parsimony, we focus on only four of these 22 variables in this paper: (1) mean daily rainfall, (2) number of days without rain, (3) mean seasonal temperature, and (4) growing degree days (GDD). All are calculated for the growing season in each country as defined by FAO.[12] These four variables are indicative of a number of different ways to measure precipitation (volume versus count) and temperature (measured temperature versus bounded count).

Fig. 4 presents the distribution of mean daily rainfall (measured in mm) during the growing season, by anonymization method and remote sensing product. In general, different anonymization methods implemented to protect privacy have only a small effect on the accuracy of measuring the volume of precipitation. Where differences occur, they tend to be deviations due to mismeasurement introduced by using Administrative boundaries (either bilinear, simple, or zonal mean methods) relative to Household bilinear. These deviations appear to be focused in the lower and center part of the distribution in all six remote sensing products. While there is not much variation between anonymization methods, there is disagreement between remote sensing products regarding the volume of precipitation in a given location. Looking across panels there are substantial differences in the distribution of rainfall as reported by each remote sensing product. CHIRPS, CPC, ARC2, and TAMSAT each report maximums in the eight to 12 mm range. By comparison, MERRA-2 reports a maximum average of 15 mm a day and ERA5 reports maximum average rainfall of nearly 42 mm. Recall, this is the mean of daily rainfall for a single growing season in a single year.

Fig. 5 further explores these differences by estimating the mean number of days without rain reported for each anonymization method by each remote sensing product in each season. Mean estimates are generated using a fractional-polynomial and graphs include 95% confidence intervals on the mean estimates. Considering the variation by anonymization method, Administrative bilinear and Administrative zonal mean clearly under count the days without rain while EA modified simple and Administrative simple tend to over count days without rain. These differences are less pronounced in products based on assimilation models. Turning to the remote sensing products themselves, CHIRPS, CPC, and ARC2 frequently report a similar number of days without rain (100–150). Similarly, MERRA-2 and ERA5 are often in agreement (40–80). TAMSAT is similar to CHIRPS, CPC, and ARC2 in the early years ($\approx$ 100), though deviates from these products in

[11] The populated pre-analysis plan is also available as a World Bank Policy Research Working Paper (Michler et al., 2021a).

[12] For more details on the definitions of growing seasons in each country, see online Appendix A.2 and Table A2.

later years (110 < 140). Measurements from CHIRPS, CPC, ARC2, and TAMSAT suggest that there are substantially more days without rain, relative to the measurements from MERRA-2 and ERA5.

In Fig. 6 we present the distribution of mean seasonal temperature (measured in °C), by anonymization method and remote sensing product. Compared to the distribution of mean daily rainfall, the figures show much tighter distributions around mean temperature, though the use of Administrative linear, Administrative simple, and Administrative zonal mean frequently result in mismeasurement. Unlike in mean daily rainfall, the deviations in temperature are almost exclusively at the lower end of the distribution. All ten anonymization methods produce essentially the same results for temperatures above 25 °C. In terms of remote sensing products, all three products tend to agree with each other, though MERRA-2 and CPC report temperatures of zero degrees, giving them long left tails.

Fig. 7 estimates the mean GDDs in a year using a fractional-polynomial and includes 95% confidence intervals on the mean estimates. As with number of days without rain, GDD represents a relative coarsening of the data by converting measured temperature into a count variable for the number of days in which temperature fell within a given range. Unlike the number of days without rain, we see no statistical differences in GDD across the ten anonymization methods or across the three remote sensing products. Confidence intervals overlap for all methods, for all remote sensing products, and in all years.

Summarizing the descriptive evidence: the use of some anonymization methods to protect privacy induces a loss of accuracy. This loss of accuracy, however, is primarily limited to the use of administrative unit or administrative area for spatial feature representation. Not surprisingly, administrative area provides the greatest degree of privacy protection but is also the least accurate in representing the precipitation and temperature experienced by the household. Reducing privacy protection by using anonymization methods that are closer to the true household location produce more accurate measurements of the weather. Mismeasurement also varies by remote sensing product, which makes intuitive sense as the products differ in their spatial resolution. Last, there is also evidence of mismeasurement in the remote sensing products themselves, with large disagreements between some products regarding daily precipitation and smaller disagreements regarding the daily temperature.

4. Analysis plan

The following analysis and the associated results were pre-specified in our pre-analysis plan (Michler et al., 2019b), which was registered with Open Science Framework (OSF). If methods, approaches, or inference criteria differ from our plan, we highlight these differences. Results arising from these deviations in our plan should be interpreted as exploratory.

4.1. Estimation

Our basic model specification follows (Deschênes and Greenstone, 2007):

$$Y_{ht} = \alpha_h + \gamma_t + \sum_j^J \beta_j f_j\left(W_{jht}\right) + u_{ht} \tag{1}$$

where Y_{ht} is our outcome variables from the LSMS-ISA-supported household surveys, described above, for household h in year t, log transformed using the inverse hyperbolic sine. We control for year fixed-effects (γ_t) and include household fixed-effects (α_h) in some specifications. The function $f_j\left(W_{jht}\right)$ represents our weather variables of interest where j represents a particular measurement of weather. Last, u_{ht} is an idiosyncratic error term clustered at the household-level.

From this general set-up, we estimate four versions of the model: two linear and two quadratic.[13] For each model, a single weather variable is considered. For the linear specification:

$$Y_{ht} = \alpha + \beta_1 W_{ht} + u_{ht} \tag{2a}$$

$$Y_{ht} = \alpha_h + \gamma_t + \beta_1 W_{ht} + u_{ht} \tag{2b}$$

For the quadratic specification:

$$Y_{ht} = \alpha + \beta_1 W_{ht} + \beta_2 W_{ht}^2 + u_{ht} \tag{3a}$$

$$Y_{ht} = \alpha_h + \gamma_t + \beta_1 W_{ht} + \beta_2 W_{ht}^2 + u_{ht} \tag{3b}$$

All of the regression models are estimated for each permutation of the data (see Table 4). This is a substantial number of regressions, given the number of variables defined (14 rainfall, eight temperature variables), the number of countries (six), the number of remote sensing products (six rainfall, three temperature), the number of anonymization methods (ten), and the number of outcomes (two). This gives us a total of 51,840 different regressions: each of our four models and two outcomes on the 540 different versions of the data. By varying both specifications and data, we seek to define a robust set of outcomes by combining the multiple analysis approach of Simonsohn et al. (2020) with the multiverse approach of Steegen et al. (2016).

4.2. Inference

In a "typical" economics paper, empirical results would be presented in a table, which would include coefficient estimates and some statistic for inference, such as standard errors, p-values, t-statistics, or confidence intervals. In our case, because of the large number of regressions that we estimate, standard modes of inference and traditional presentations of results are not appropriate. Instead, per our pre-analysis plan, we rely on a series of methods and criteria to make inference, evaluate the results, and present our findings.[14]

As no formal statistical test exists to compare results across models, we develop three heuristics that allow us to describe similarities and differences in our results. Before describing these heuristics, it is useful to reflect on what sort of characteristics a heuristic would need to be useful for our purposes (i.e., comparing across tens of thousands of model-data combinations). First, some weather metrics that we test are likely to be positively correlated with outcomes (mean rainfall) while others are likely to be negatively correlated (days without rain). So, a heuristic should be agnostic about the sign of the coefficient. Second, our prior is that weather is significantly correlated with outcomes, regardless of direction. This maintained assumption is based on the frequency with which weather is used in the economics literature to predict all sorts of outcomes, from crop production to migration to economic growth. As such, one would want a heuristic that is able to determine when a weather metric is significantly correlated with

[13] In our pre-analysis plan we defined two additional models that include measured inputs (fertilizer, labor, pesticide, herbicide, and irrigation). However, we find that controlling for inputs has no discernible effect on results, relative to the household fixed effects model and so we exclude these results from this paper. The populated pre-analysis plan on arXiv.org and through the World Bank contain all of these results (Michler et al., 2021b,a).

[14] As specified in our pre-analysis plan, we intended to examine the CDFs of coefficient estimates, following Sala-i-Martin (1997b,a). However, using this approach in our context did not yield informative results. As such, we instead graph coefficients and confidence intervals ordered by the size of the coefficient estimate in specification charts. While not the same as the CDFs of coefficients in Sala-i-Martin (1997a,b), the graphs communicate roughly the same information and are more appropriate for the variation in metrics, data products, anonymization methods, and so on, which are relevant for this analysis.

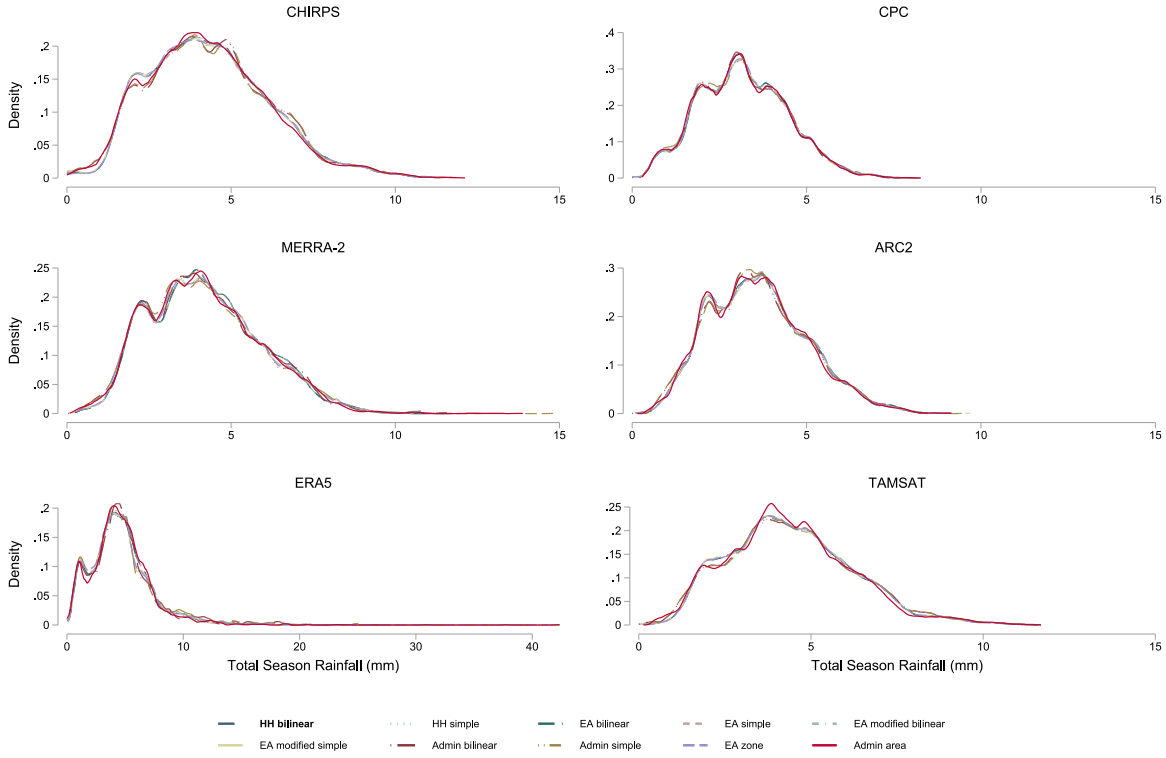

Fig. 4. Distribution of mean daily rainfall, by anonymization method and remote sensing source.
Note: The figure presents rainfall distributions pooled across all countries and years, disaggregated by remote sensing source. Each line (anonymization method) in each panel is constructed using all 33,738 household-year observations. Variation in lines does not come from variation in the household data that is paired with the remote sensing data. Rather, variation in lines within a panel is solely due to differences in the grid cell in which the anonymization method locates the household. Variation in lines across panels is solely due to differences in the value of precipitation reported by the remote sensing source.

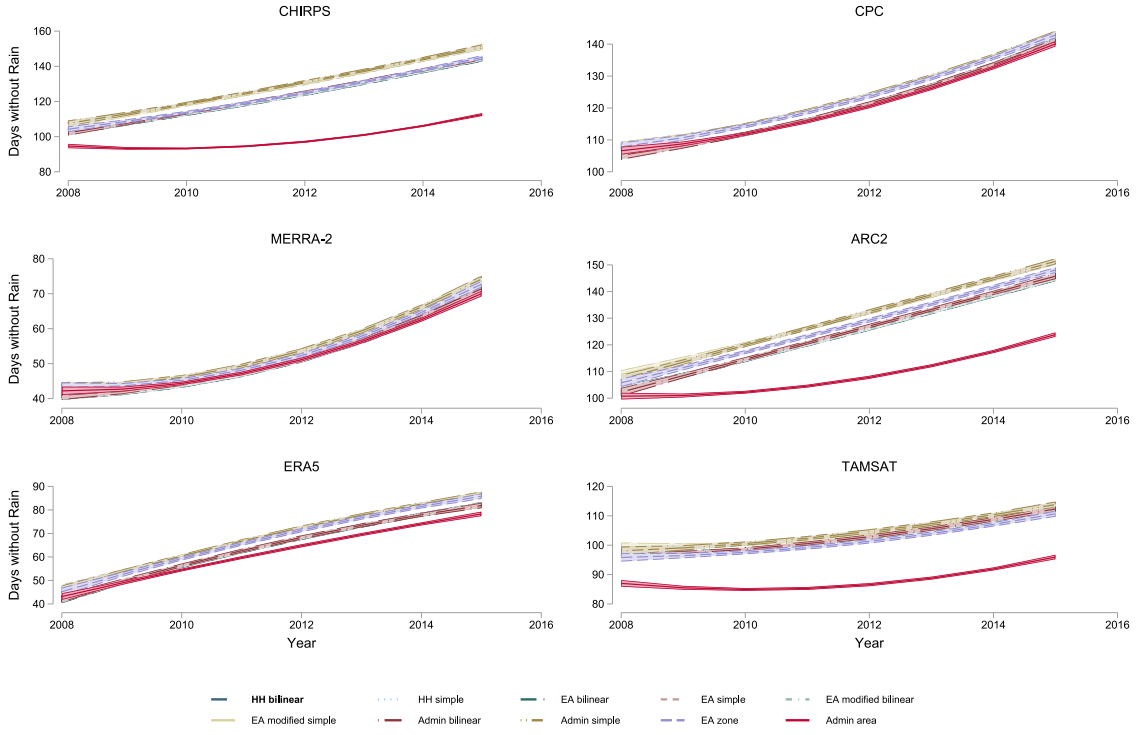

Fig. 5. Prediction of Mean number of days without rain, by anonymization method and remote sensing source.
Note: The figure presents the mean number of days without rain (<1 mm) in a year, pooled across all countries, disaggregated by remote sensing source. Prediction is made via Fractional-Polynomial, with 95% confidence intervals represented by the shaded area. Each line (anonymization method) in each panel is constructed using all 33,738 household-year observations. Variation in lines does not come from variation in the household data that is paired with the remote sensing data. Rather, variation in lines within a panel is solely due to differences in the grid cell in which the anonymization method locates the household. Variation in lines across panels is solely due to differences in the number of days without rain reported by the remote sensing source.

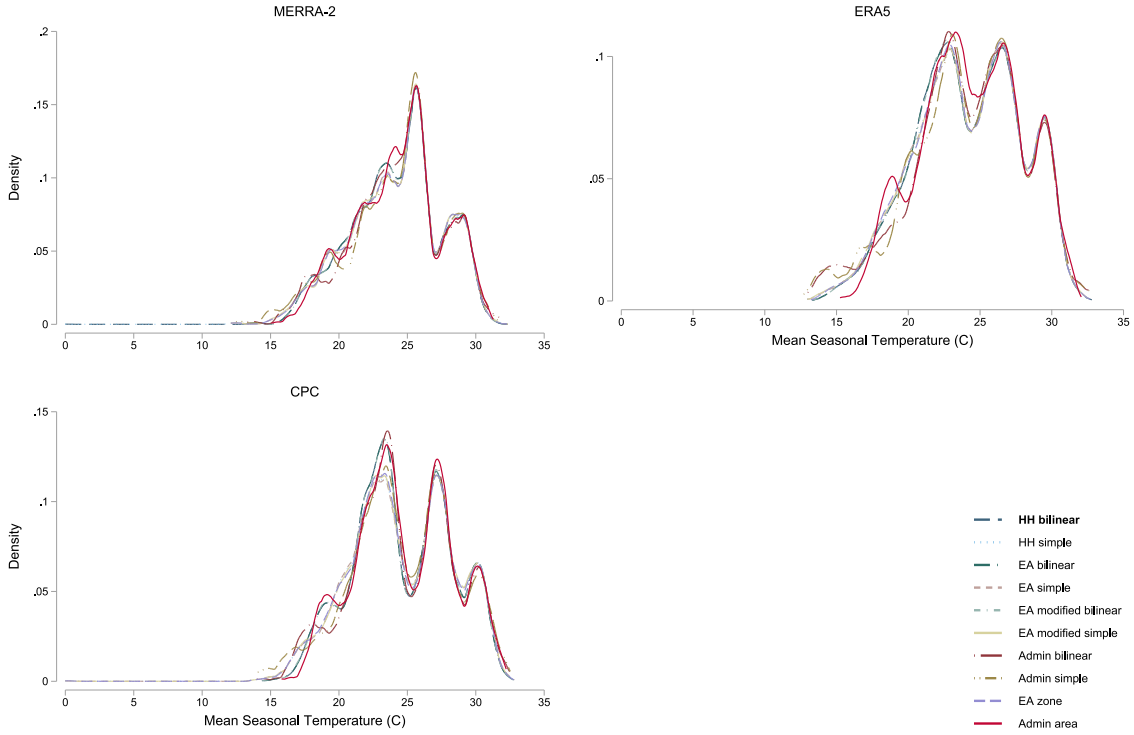

Fig. 6. Distribution of mean seasonal temperature, by anonymization method and remote sensing source.
Note: The figure presents temperature distributions pooled across all countries and years, disaggregated by remote sensing source. Each line (anonymization method) in each panel is constructed using all 33,738 household-year observations. Variation in lines does not come from variation in the household data that is paired with the remote sensing data. Rather, variation in lines within a panel is solely due to differences in the grid cell in which the anonymization method locates the household. Variation in lines across panels is solely due to differences in the value of temperature reported by the remote sensing source.

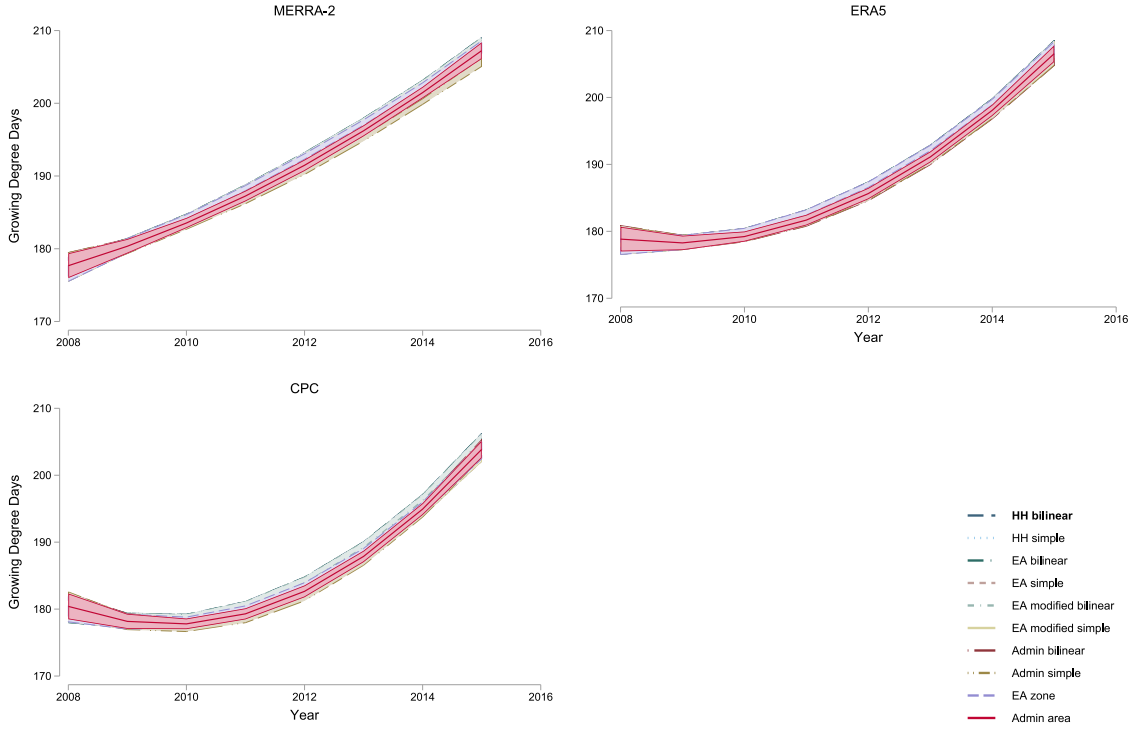

Fig. 7. Prediction of mean number of growing degree days, by anonymization method and remote sensing source.
Note: The figure presents the mean number of growing degree days (GDD) in a year, pooled across all countries, disaggregated by remote sensing source. Prediction is made via Fractional-Polynomial, with 95% confidence intervals represented by the shaded area. Each line (anonymization method) in each panel is constructed using all 33,738 household-year observations. Variation in lines does not come from variation in the household data that is paired with the remote sensing data. Rather, variation in lines within a panel is solely due to differences in the grid cell in which the anonymization method locates the household. Variation in lines across panels is solely due to differences in the value of temperature reported by the remote sensing source.

outcomes and when it is not. Last, and in line with our prior, we expect weather to reduce the amount of unexplained variance in a model, all else being equal. So, one would want a heuristic that can measure the amount of unexplained variance in the model after controlling for weather.

With these three characteristics in mind, we adopt three general metrics to evaluate our results and two methods to test differences between these metrics. The three metrics are (1) mean log likelihood values, (2) share of coefficient p-values significant at standard levels (0.01, 0.05, and 0.10), and (3) coefficient size with 95% confidence intervals. To compare our metrics across regressions, we apply two tests:

1. *Weak difference test*: the value of a result (either mean log likelihood, share of significant p-values, or coefficients) from one regression lies outside the 95% confidence interval on the value of a result from a competing regression. The confidence intervals *can* overlap.
2. *Strong difference test*: the 95% confidence interval on the value of a result (either mean log likelihood, share of significant p-values, or coefficients) from one regression lies outside the 95% confidence interval on the value of a result from a competing regression. The confidence intervals *cannot* overlap.

Our approach builds on the extreme bounds approach to assessing differences in estimates from Levine and Renelt (1992) and the graphical methods to visualize these differences in Sala-i-Martin (1997a,b).

While the three metrics are formal statistics, our weak and strong tests are not and we do not treat them that way. Rather, we use the combination of metrics and informal tests as heuristics in evaluating the loss of accuracy (mismeasurement) induced by anonymization methods used to protect participant privacy. All comparisons of one obfuscation/metric/source combination are made relative to the Household bilinear/metric/source combination. Our heuristics do not allow us to make claims regarding a formal definition of statistical accuracy, such as the expected squared-error loss in Abowd and Schmutte (2019). Rather, we quantify the significance and magnitude of measurement error by comparing results from one anonymization method with results from Household bilinear always bearing in mind that, for a given metric and country, if there was no measurement error induced by anonymization method, then the results from our tens of thousands of regressions would be exactly the same regardless of the obfuscation/source combination.

An important caveat to bear in mind with respect to our results, in particular all of the results focused on p-values, is that the significance of a point estimate does not imply that the model is correctly specified, that the point estimate is agronomically meaningful, or that the point estimate has the correct sign. These results and the associated figures simply allow us to visualize the variability in the number of significant coefficients across these specifications of interest. And any variability in results is a sign that obfuscation/source combinations provide different measures of weather and measurement error thus exists.

5. Results

We present results in a series of figures, which allow us to evaluate the significance, magnitude, and general trends in how methods undertaken to preserve privacy affect accuracy. We do this due to the large number of regressions and estimated values produced in our analyses which make standard presentations of empirical results inappropriate.

To examine the impact that different obfuscation procedures have on agricultural productivity, we pool the results from the 51,840 regressions and then divide the pool into ten bins, one for each anonymization method. In order to evaluate these outcomes, following the heuristics for inference discussed above, we then calculate descriptive statistics for each bin of results. These include the mean log likelihood value and the share of coefficients (β_1) with p-values of $p > 0.90, p > 0.95$ or $p > 0.99$. For each of these values, we calculate the 95% confidence interval on the mean. We then compare mean log likelihood values or the share of $p > 0.95$ across all ten anonymization methods and use the 95% confidence interval on the mean to evaluate differences using our weak and strong test criteria. Last, we use specification charts to examine the actual regression coefficients and estimated confidence intervals for a subset of regressions.

5.1. Log likelihood

We use specification charts to examine log likelihood values across the ten types of anonymization methods. The value of the log likelihood function is a measure of explained variance in the model, so models with more accurate data (less measurement error) are likely to have a smaller amount of variance left unexplained. Fig. 8 shows the mean log likelihood and the 95% confidence interval on the mean by anonymization method. We further disaggregate results by model specification, as a model with fixed effects will have a different log likelihood value than a model without fixed effects. The top panels of Fig. 8 displays results from model specifications (2a) and (3a), which are the linear and quadratic models without household or year fixed effects. The bottom panel displays results from model specifications (2b) and (3b), which include household and year fixed effects. Within each specification chart, at the top of each "column" is the mean log likelihood and the 95% confidence interval on the mean for the set of 1296 regressions run. Below, markers on the chart indicate the anonymization method associated with the statistics. Household bilinear, which represents the true household coordinates, is highlighted in the figures with an orange marker for easier reference.

Consider first the specification chart in the top panel which include only weather as an explanatory variable. Mean log likelihood values are not different across anonymization method within model specification (2a). The mean log likelihood value for any one anonymization method fails to pass even our weak difference test when compared to Household bilinear. Similarly, when comparing across anonymization methods within model specification (3a), no mean log likelihood is weakly different from Household bilinear.

We conduct the same exercise for results presented in the bottom panels from model specifications that include fixed effects. As with the top panel, the mean log likelihood value for any one anonymization method is not even weakly different from Household bilinear. Our heuristic fails to identify significant differences within any model specification. Based on this, we conclude that remote sensing weather data from any one anonymization method does not explain a substantially larger amount of the variance in our outcome variables relative to the true household coordinates.

Despite the failure to identify differences in anonymization method, based on either the strong or weak criteria, the pattern of which anonymization methods result in the largest log likelihood values is remarkably consistent. Household bilinear, EA bilinear, and EA modified bilinear always make up three of the top four models. Recall that the bilinear method computes the distance weighted average of values at the four nearest cell centers. Thus, unlike the simple extraction method, the bilinear method accounts for the point location within the arbitrary cell boundaries of the gridded data product. This approach seems to produce slightly better results than the simple extraction method for points or the EA zone of uncertainty. Administrative area appears to be too large of an area to produce strong results, as using Administrative area, regardless of point or polygon representation, tends to produce the smallest log likelihood values. While the pattern is consistent, it is important to recall that differences between each spatial anonymization method and Household bilinear is not substantial enough to pass even our weak test, and we fail to identify significant differences across methods.

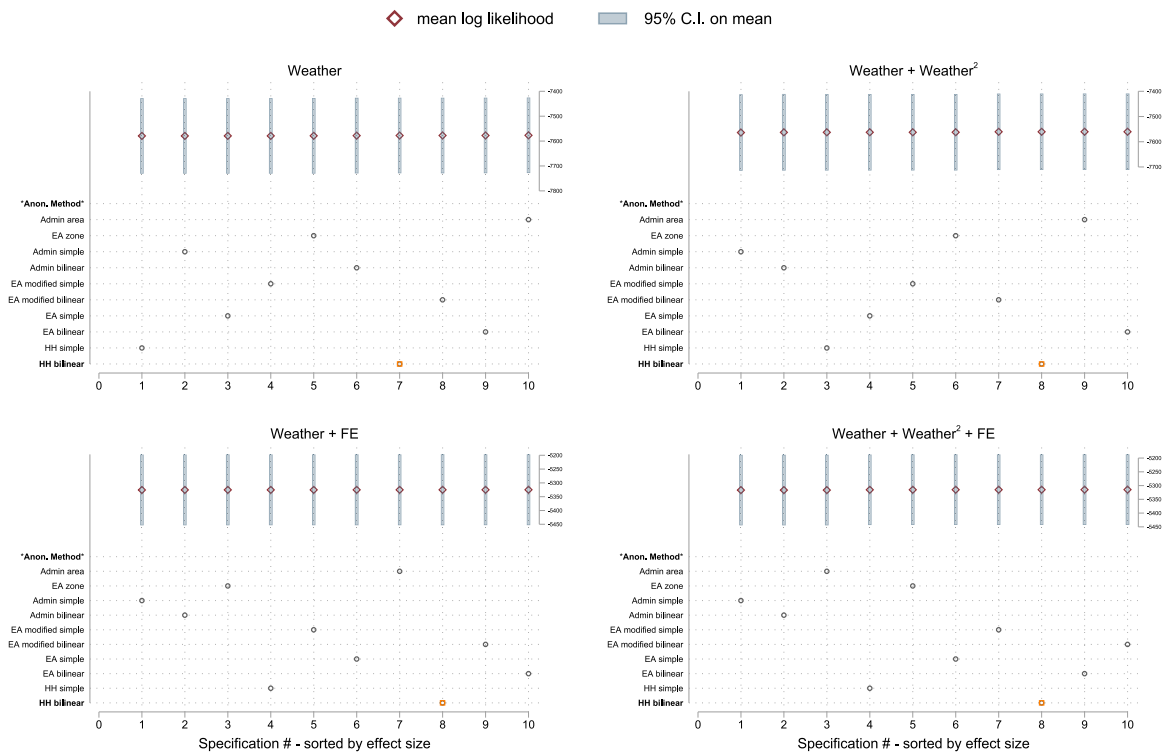

Fig. 8. Mean log likelihood, by extraction and model.
Note: The figure presents the mean log likelihood, by anonymization method and model specification, aggregated over country, weather metric, remote sensing source, and outcome variable. The figure is derived from the results of all 51,840 regressions, with each panel summarizing the results of 12,960 regressions. Each column in each panel summarizes the results of 1296 regressions, which are for each specification model and each anonymization method. Orange diamonds identify results using the true household coordinates using the bilinear extraction. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

5.2. *p*-values

We next consider if different anonymization methods produce substantially different counts of significant coefficients. Although while examining log likelihood values we disaggregated each bin of regression results by model specification, when examining *p*-values we disaggregate by whether the remote sensing data is rainfall or temperature. Fig. 9 presents the share of significant coefficient estimates for three standard *p*-values: for $p > 0.90, p > 0.95$ or $p > 0.99$. To these bars we add the 95% confidence interval on the mean number of significant coefficients. The top panel presents results from precipitation products while the bottom panel presents results from temperature products. Each bar and confidence interval in the rainfall panel is based on 4032 regressions while each bar and confidence interval in the temperature panel is based on 1152 regressions. To facilitate comparison, we draw red lines to designate the top and bottom of the confidence interval on the mean for the Household bilinear method, which are the actual household coordinates.

A quick, visual inspection of the results in the top panel of Fig. 9 does not reveal many, if any, differences across anonymization method. Comparing numerical values for the share of significant coefficients from Household bilinear to the 95% confidence interval on the mean of any other extraction reveals that there are no comparisons that are strongly different from each other. There is only one weak difference, that of Administrative area, which produces slightly more significant *p*-values than those produced by data matched to the true household coordinates. Similarly, the results in the lower panel on temperature look fairly uniform across anonymization methods. No pairwise comparisons to Household bilinear are strongly different or weakly different.

However, there is a possibility of heterogeneity across or within countries. As such, we next consider this same metric, disaggregated by country. Figs. 10 and 11 present different anonymization methods across all rainfall and temperature metrics, for each of the six countries. Now that we have divided the results by anonymization method, rainfall/temperature, and country, each bar represents the share of significant coefficients from 672 regressions for rainfall and 192 for temperature. We simplify the graph by only presenting the share of coefficients with $p > 0.95$.

We see some variation within countries based on anonymization method. While no anonymization method is strongly different from Household bilinear, in Ethiopia, Niger, Nigeria, and Uganda, there are some methods that are weakly different. In all cases, these differences are from using administrative unit or area. In Ethiopia, Administrative simple and Administrative bilinear are weakly different from Household bilinear. In Niger, both Administrative bilinear and Administrative zonal mean are weakly different from Household bilinear while in Nigeria, Administrative zonal mean is weakly different from Household bilinear. In Uganda, Administrative simple is weakly different from Household bilinear. There are no significant differences in Malawi or Tanzania. That all significant differences are associated with Administrative unit or area suggests that this approach to privacy protection does come at the cost of some data accuracy, though again the differences are only weak and are not present in all countries.

Considering temperature, the evidence for differences in anonymization method is noisy (larger confidence intervals) relative to rainfall. As a result, there is no apparent pattern of one anonymization method differing from Household bilinear. One exception to this is the case of Ethiopia, in which there are weak differences between Household bilinear and Household simple, EA simple, EA modified simple, EA zone of uncertainty, and Administrative area. But, no other countries show any differences, weak or strong, between Household bilinear and any anonymization method.

As with our examination of log likelihood values, the preponderance of evidence on *p*-values implies that different anonymization methods used to protect privacy do not introduce substantial mismeasurement

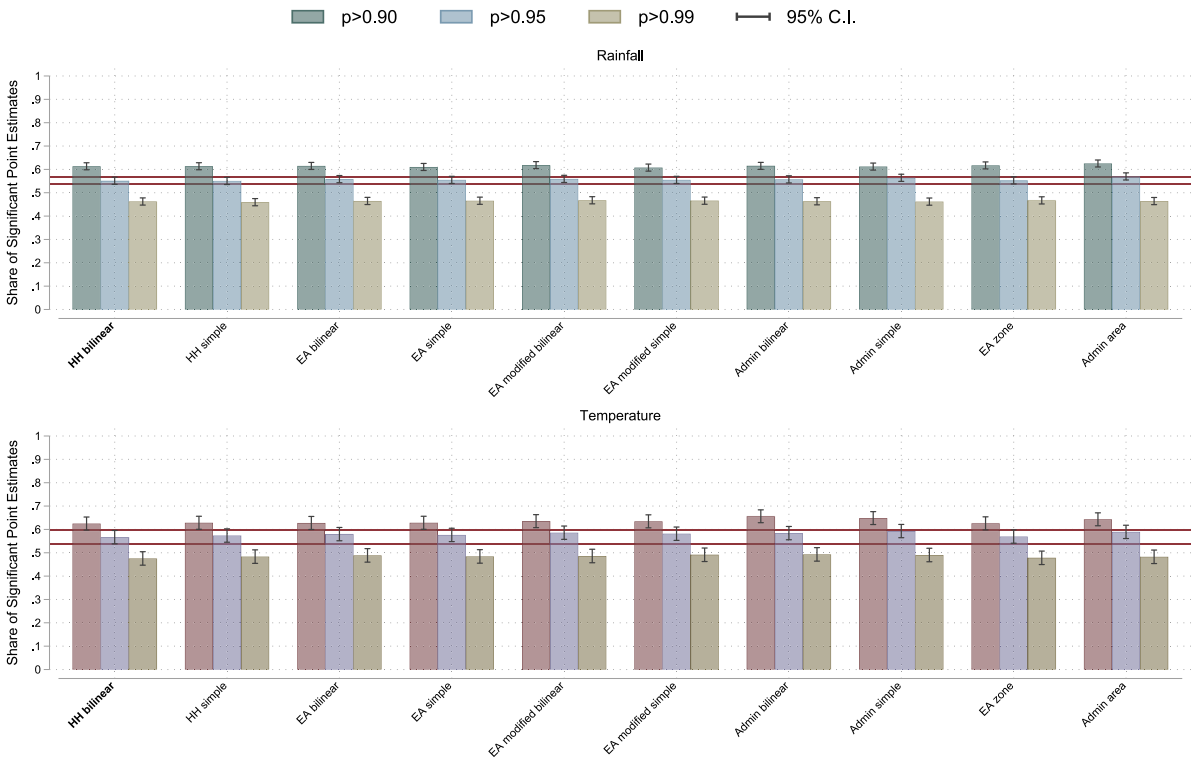

Fig. 9. *p*-values of rainfall and temperature, by anonymization method.
Note: The figure displays the share of coefficients on the rainfall and temperature variables that are statistically significant from each anonymization method, aggregated over country, weather metric, remote sensing source, outcome variable, and specification. The northern panel presents rainfall while the southern panel presents temperature. The data summarized in the northern panel includes 40,320 regressions, with each column including 4032 regressions. The data summarized in the southern panel includes 11,520 regressions, with each column including 1152 regressions. Red lines designate the top and bottom of the confidence interval on the mean for the true household coordinates using the bilinear extraction. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

into the analysis. There are some weak differences, as with log likelihoods, when comparing administrative unit or area to Household bilinear, particularly when using precipitation data. There are also some differences between Household bilinear and other methods when results are disaggregated by country. Again, these differences tend to be weak and exist only when we compare administrative unit or area to Household bilinear.[15]

5.3. Coefficients

In order to be able to examine individual regression coefficients, we first must narrow our focus to a subset of the 51,840 results. To do this, we consider four weather metrics: mean daily rainfall, number of days without rain, mean seasonal temperature, and growing degree days.[16] We also focus on two models: weather only and weather with year and household fixed effects.[17] Similar to the specification charts for log likelihood, labels identifying characteristics of the results are presented at the bottom of the specification chart. Unlike the log likelihood charts, we now present coefficients and confidence intervals for single regressions–120 results per rainfall metric per country and 60 results per temperature metric per country–and not means of aggregated results and confidence intervals on the mean. Thus we present specific coefficient estimates from 4320 regressions. In the following discussion, the term significance defines a point estimate with $p > 0.95$. Household bilinear, which represents the true household coordinates, is highlighted in the figures with an orange marker for easier reference.

Figs. 12 through 17 present specification charts for coefficients and confidence intervals on mean daily rainfall and the number of days without rain by country. A number of patterns are immediately obvious. Results vary systematically by country, model, remote sensing product, and dependent variable. What is not clear is how results vary by anonymization method. In many countries and in both models, markers indicating remote sensing product or dependent variable tend to cluster within a specification chart, suggesting a pattern to results. Consider, as an example, in Ethiopia rainfall tends to be more strongly correlated (measured by a large absolute value of coefficient size) with yield than with value of harvest. No pattern of clustering exists for anonymization method, regardless of country, model, remote sensing product, or weather metric. The markers for anonymization method appear as random noise in each specification chart, suggesting that, relative to other sources of variation, anonymization method does not have a systematic impact on coefficient size or significance.

Turning to temperature, results regarding the impact of anonymization method are qualitatively similar to rainfall. In Figs. 18 through 23, markers for anonymization method appear to be nearly random while markers for remote sensing weather product and dependent variable cluster depending on the country, model, and temperature metric. As with rainfall, variation from country, model, remote sensing product, or

[15] There are patterns to the variation across countries with respect to the share of significant *p*-values. The pattern is not the result of mismeasurement but is interesting to note for the discussion of cross-country differences in weather's relationship to agricultural productivity. Michler et al. (2021a) explores in more detail these relationships and their implications for integrating remote sensing weather data with household survey data.

[16] Results and conclusions do not change in a meaningful way if we use any of the other 18 weather metrics instead of these four. These four were chosen to provide evidence from different ways to measure precipitation (volume versus count) and temperature (actual temperature versus bounded count). Additional results for weather shocks can be found in online Appendix C. Complete results for all 22 weather metrics are available in our populated pre-analysis plan (Michler et al., 2021a).

[17] Results and conclusions do not change in a meaningful way if we instead use the quadratic specifications. Results for the quadratic specifications are in online Appendix C.

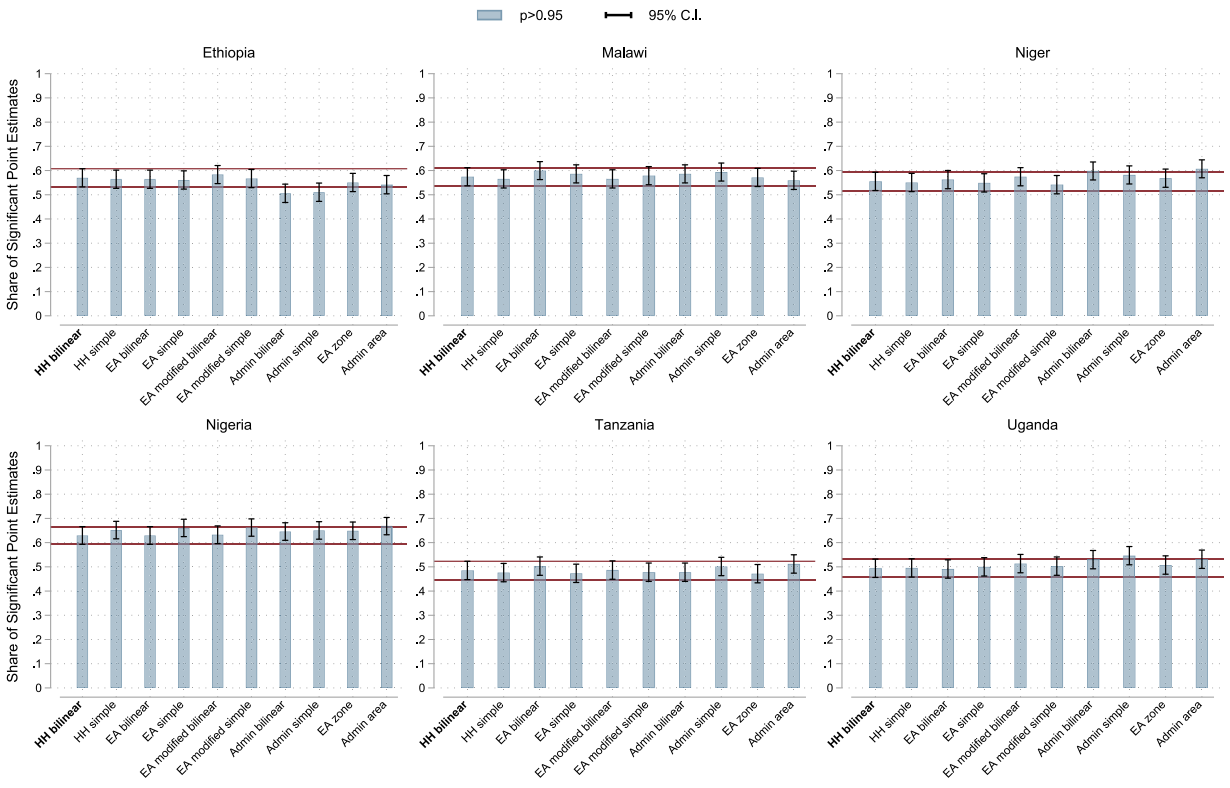

Fig. 10. *p*-values of rainfall, by country and anonymization method.
Note: The figure displays the share of coefficients on the rainfall variables that are statistically significant from each anonymization method for each country, aggregated over weather metric, remote sensing source, outcome variable, and specification. The figure presents results from a total of 40,320 regressions. Each country includes results from 6720 regressions and thus each column is based on 672 regressions. Red lines designate the top and bottom of the confidence interval on the mean for the true household coordinates using the bilinear extraction. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

weather metric appears to be more of a factor in determining coefficient sign, size, and significance than anonymization method.

Taken together, the preponderance of evidence from all of our 51,840 regressions regarding our heuristics lead us to conclude that, generally, there is no clear evidence that different SDL methods implemented to preserve privacy of farms or households have substantially different impacts on estimates of agricultural productivity. One exception to this is that Administrative measurements produce some differences, though relatively small discrepancies, in the share of significant *p*-values. As in the descriptive statistics, we find evidence that while anonymization methods that rely on administrative unit or area provide the greatest degree of privacy protection they result in losses in accuracy for measurement of precipitation experienced by the household and correspondingly mismeasure the relationship between weather and agricultural productivity. Outside of the use of administrative unit or area, however, our findings suggest that any measurement error which may arise from the use of different anonymization methods does not substantially affect estimates. When researchers use publicly available data with GPS information obfuscated using methods similar to those in the DHS and LSMS, they should feel confident that matching those coordinates with remote sensing data will not introduce substantial measurement error into the analysis.

5.4. Ancillary results

In Figs. 12 through 23 we fail to observe patterns in coefficients as a function of anonymization method. However, there are strong patterns based on country, specification, remote sensing product, and dependent variable. While the focus of this paper is on the effect of measurement error introduced by anonymization method, digging further into the specification charts reveals intriguing ancillary results based on these other sources of variation.

In terms of heterogeneity across countries, results in Ethiopia and Malawi are quite consistent when examining models with only the weather metric on the right hand side. Mean daily rainfall is either positively correlated with outcomes or it is not significant. Conversely, the number of days without rain is a either negatively correlated with outcomes or it is not significant. This pattern persists in Niger and Nigeria, though precipitation measured by MERRA-2 in Niger and ERA5 in Nigeria produces coefficients with opposite signs (negative for mean rain and positive for no rain days). In Tanzania and Uganda, there is little consistency across regressions, with about an equal number of regressions reporting positive and negative coefficients. In Tanzania, this appears to be driven by the choice of dependent variable (more rain reduces the value of harvest but increases yield) while in Uganda it appears to be driven by the choice of remote sensing product (for ARC2 and TAMSAT more rain is negatively correlated with outcomes).

The primary impact of including fixed effects in the regressions is to weaken the correlation between rainfall and outcomes. In Ethiopia, without fixed effects, rainfall is always significantly correlated with outcomes but by including fixed effects rainfall is no longer significantly related to outcomes in a majority of regressions. Results are similar in Malawi, Niger, and Nigeria, suggesting that once time-invariant household unobservables are controlled for, rainfall matters little to agricultural productivity. Tanzania and Uganda again prove to be outliers. Where without fixed effects, rainfall could be both positively and negatively correlated with outcomes, by including fixed effects results in these countries become much more consistent. In Tanzania rainfall tends to be uncorrelated with value of harvest but is consistently significantly correlated with yield. In Uganda, the results are the opposite, with rainfall significantly correlated with value of harvest but not yield.

Focusing on the temperature results, mean seasonal temperature is either negatively correlated with outcomes or not significant in

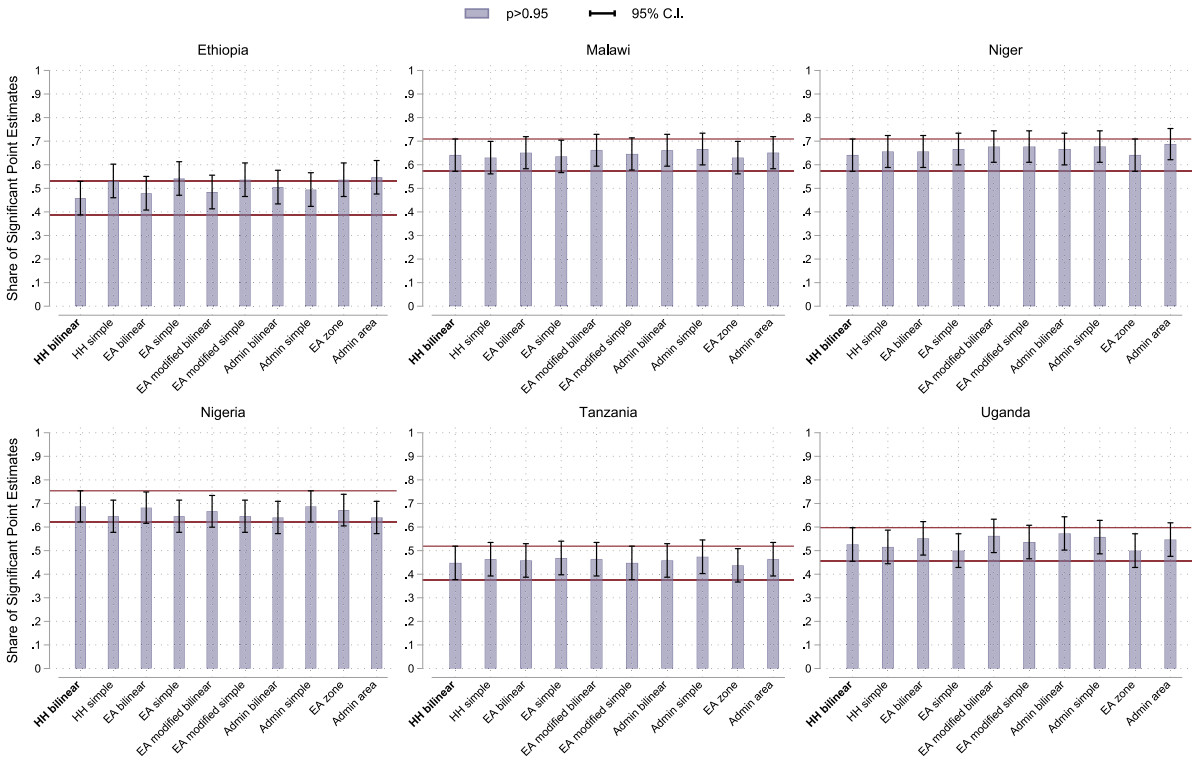

Fig. 11. *p*-values of temperature, by country and anonymization method.
Note: The figure displays the share of coefficients on the temperature variables that are statistically significant from each anonymization method for each country, aggregated over weather metric, remote sensing source, outcome variable, and specification. The figure presents results from a total of 11,520 regressions. Each country includes results from 1920 regressions and thus each column is based on 192 regressions. Red lines designate the top and bottom of the confidence interval on the mean for the true household coordinates using the bilinear extraction. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

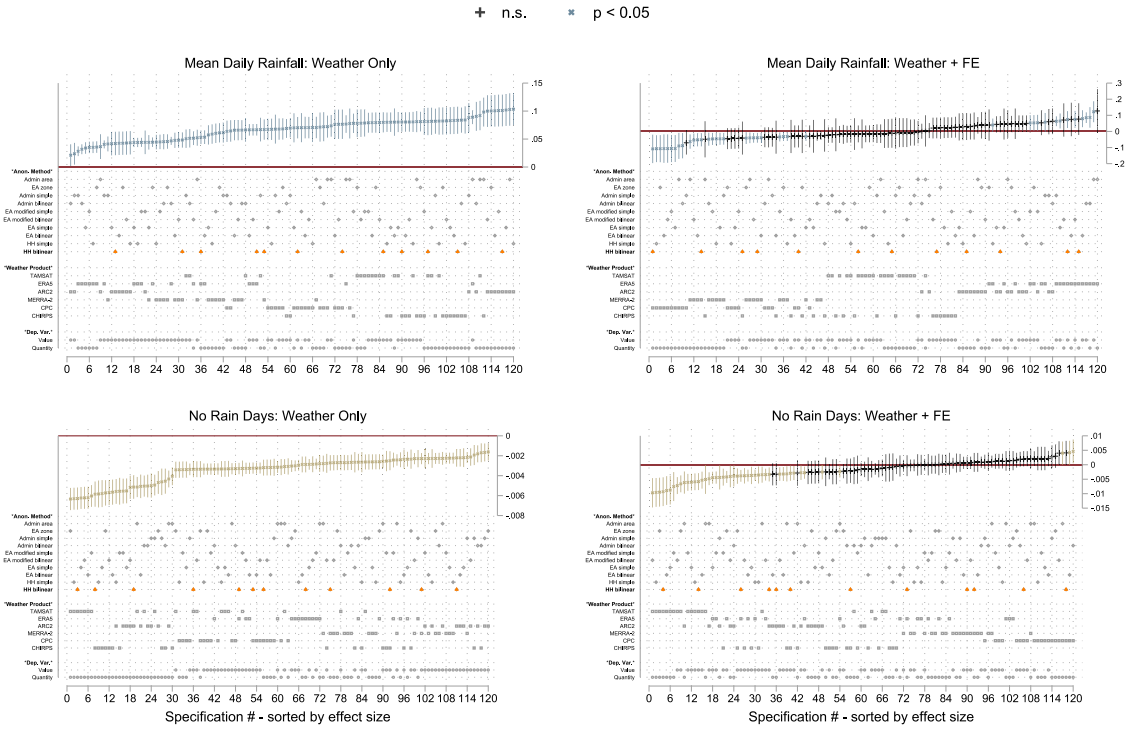

Fig. 12. Specification curve for rainfall variables in Ethiopia.
Note: The figure presents specification curves where each panel presents results from a different model. Each panel includes 120 regressions, where each column represents a single regression. Significant and non-significant coefficients are designated at the top of the figure. Orange diamonds identify results from the true household coordinates using the bilinear extraction. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

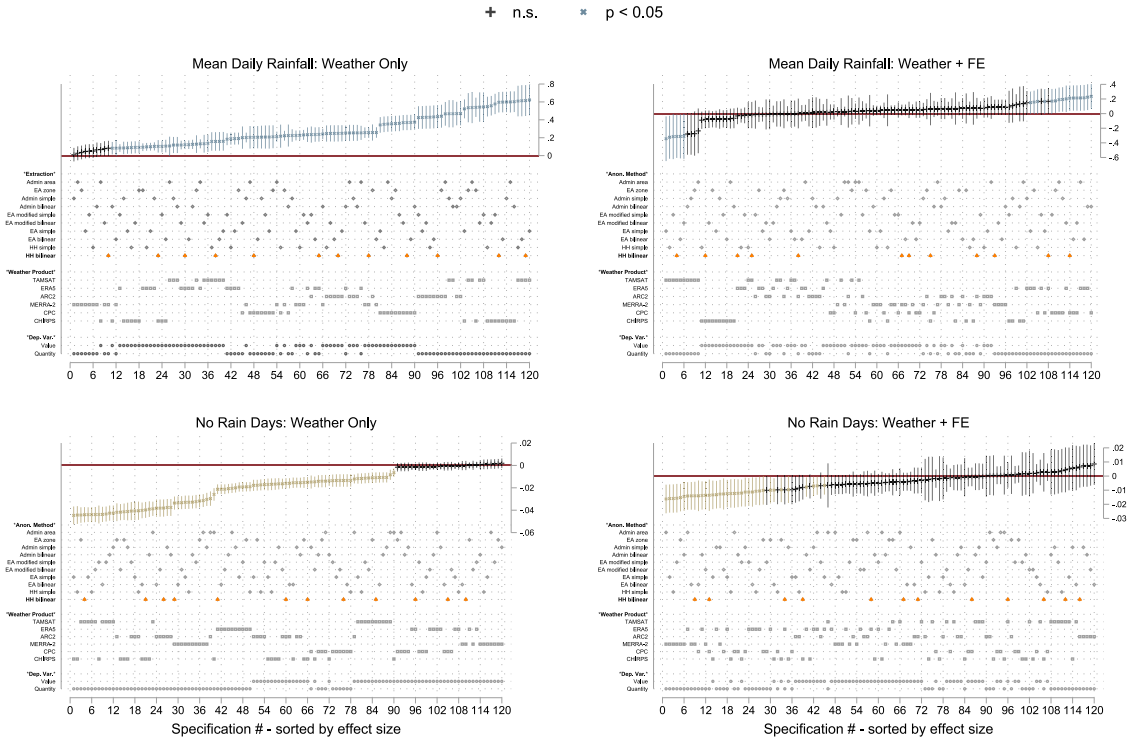

Fig. 13. Specification curve for rainfall variables in Malawi.
Note: The figure presents specification curves where each panel presents results from a different model. Each panel includes 120 regressions, where each column represents a single regression. Significant and non-significant coefficients are designated at the top of the figure. Orange diamonds identify results from the true household coordinates using the bilinear extraction. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

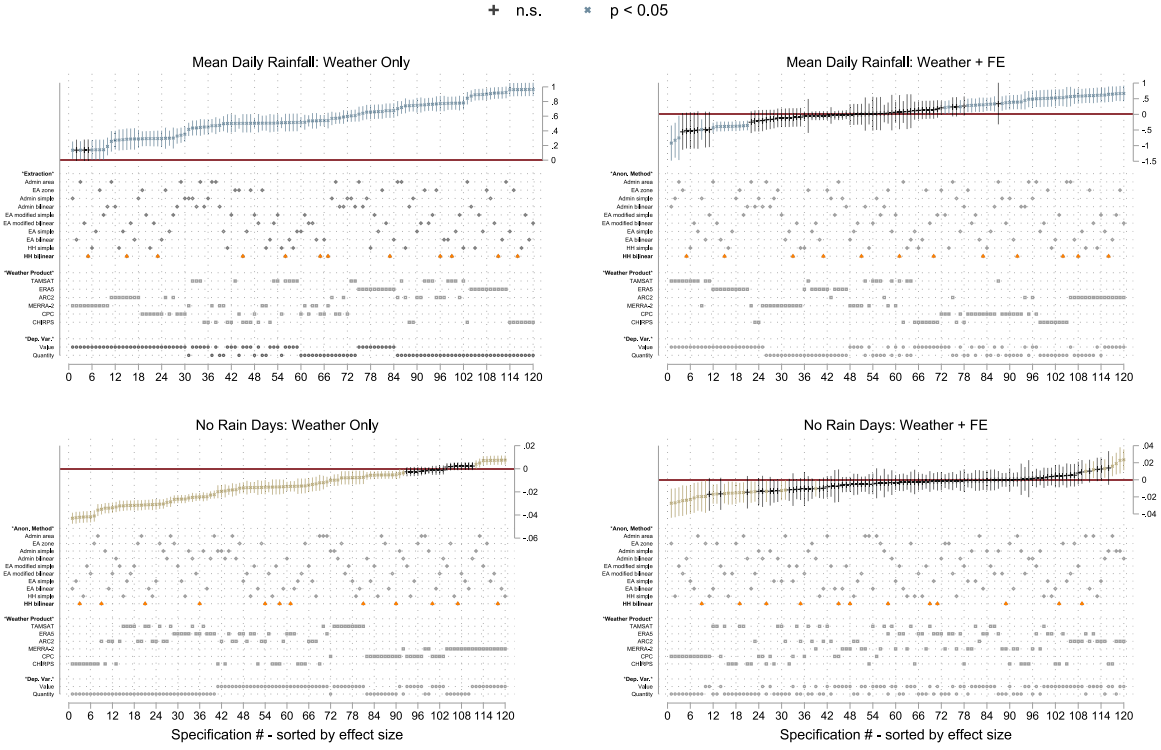

Fig. 14. Specification curve for rainfall variables in Niger.
Note: The figure presents specification curves where each panel presents results from a different model. Each panel includes 120 regressions, where each column represents a single regression. Significant and non-significant coefficients are designated at the top of the figure. Orange diamonds identify results from the true household coordinates using the bilinear extraction. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

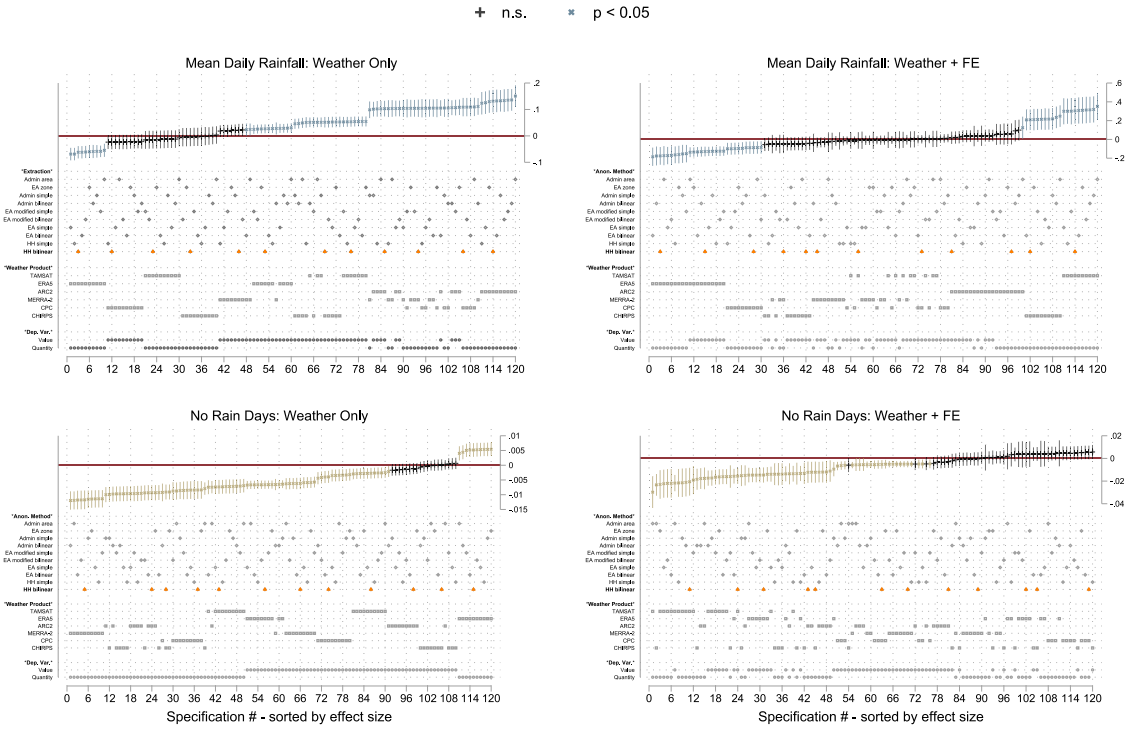

Fig. 15. Specification curve for rainfall variables in Nigeria.
Note: The figure presents specification curves where each panel presents results from a different model. Each panel includes 120 regressions, where each column represents a single regression. Significant and non-significant coefficients are designated at the top of the figure. Orange diamonds identify results from the true household coordinates using the bilinear extraction. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

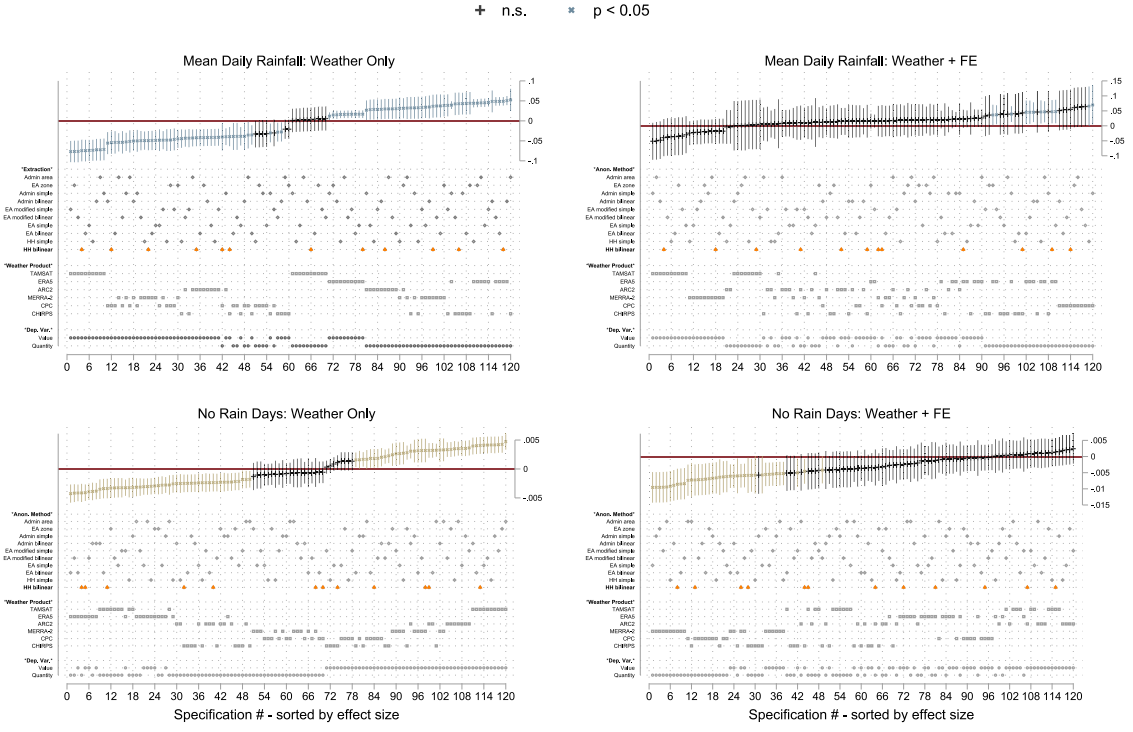

Fig. 16. Specification curve for rainfall variables in Tanzania.
Note: The figure presents specification curves where each panel presents results from a different model. Each panel includes 120 regressions, where each column represents a single regression. Significant and non-significant coefficients are designated at the top of the figure. Orange diamonds identify results from the true household coordinates using the bilinear extraction. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

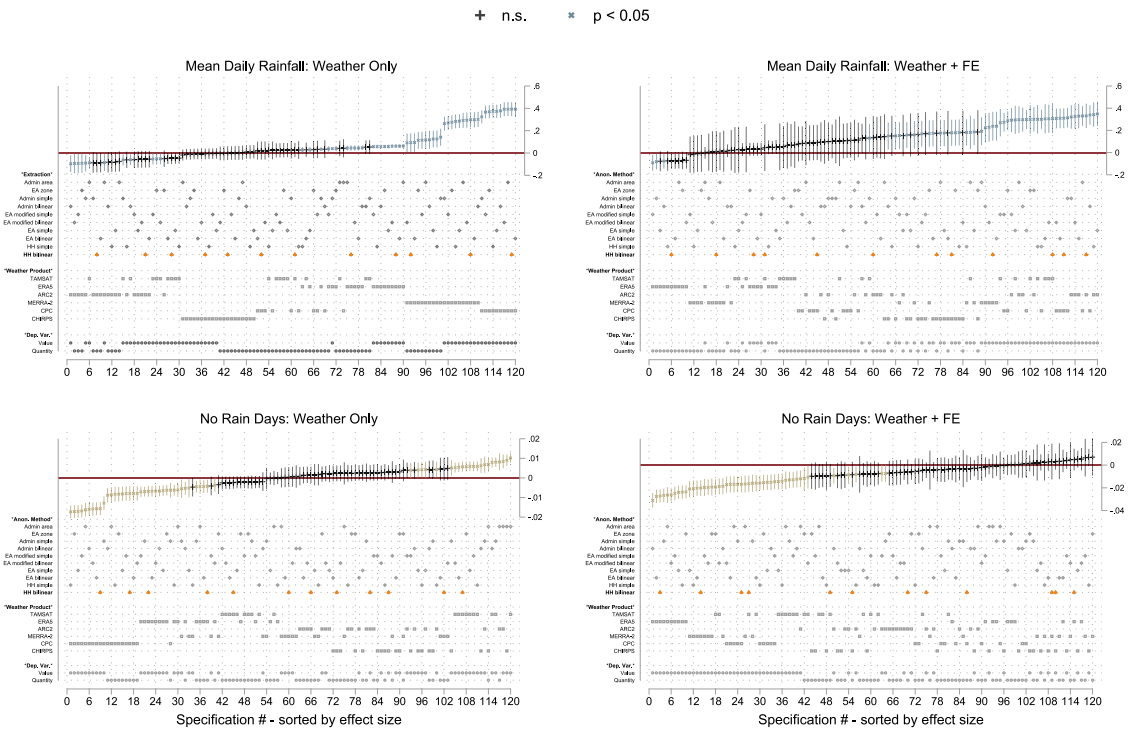

Fig. 17. Specification curve for rainfall variables in Uganda.
Note: The figure presents specification curves where each panel presents results from a different model. Each panel includes 120 regressions, where each column represents a single regression. Significant and non-significant coefficients are designated at the top of the figure. Orange diamonds identify results from the true household coordinates using the bilinear extraction. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

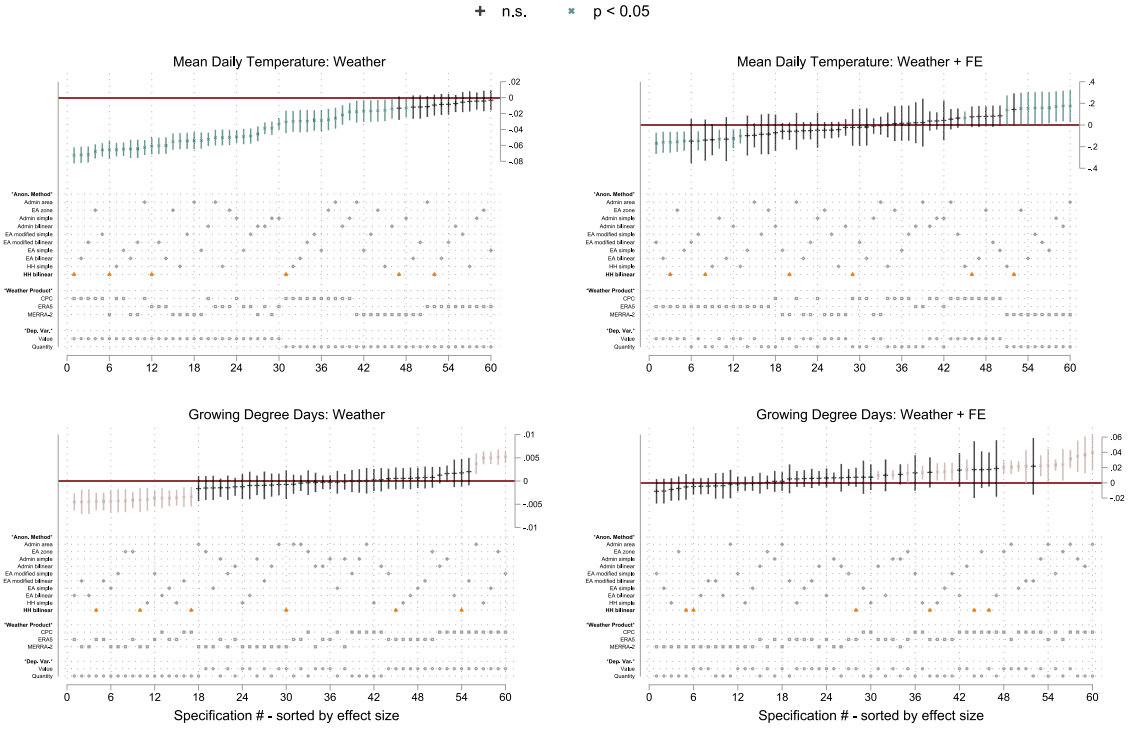

Fig. 18. Specification curve for temperature variables in Ethiopia.
Note: The figure presents specification curves where each panel presents results from a different model. Each panel includes 60 regressions, where each column represents a single regression. Significant and non-significant coefficients are designated at the top of the figure. Orange diamonds identify results from the true household coordinates using the bilinear extraction. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

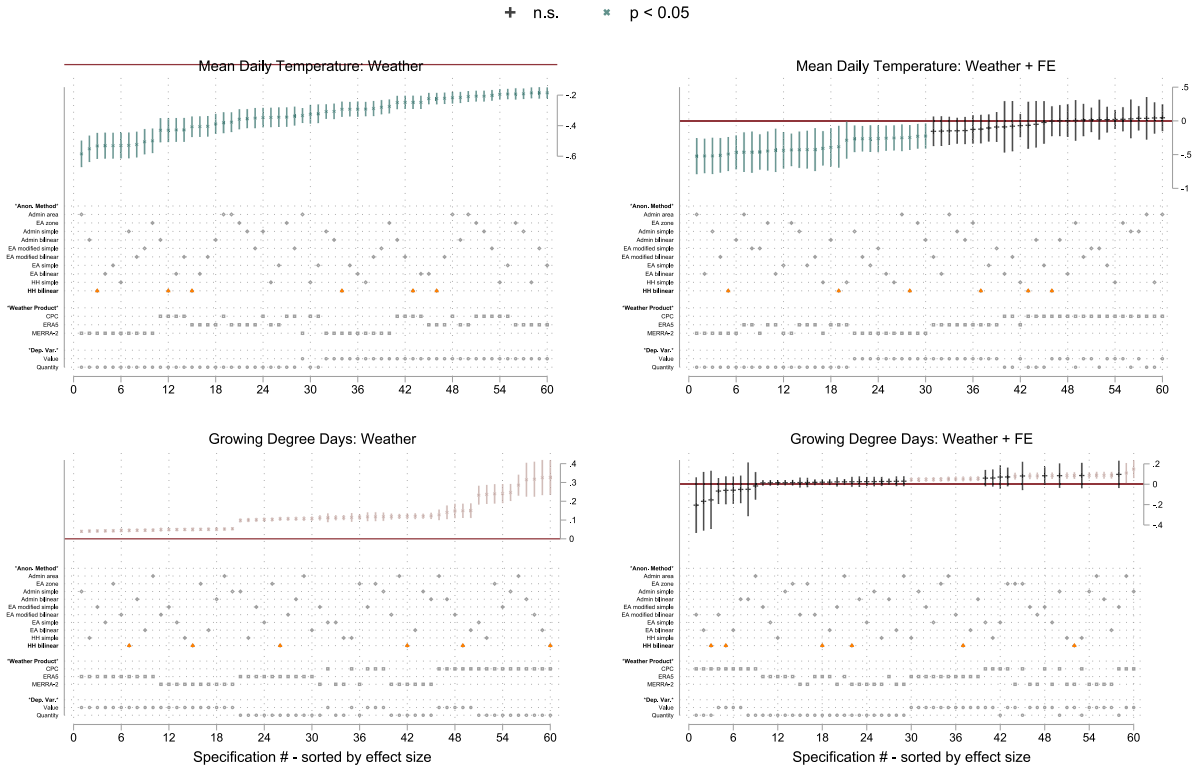

Fig. 19. Specification curve for temperature variables in Malawi.
Note: The figure presents specification curves where each panel presents results from a different model. Each panel includes 60 regressions, where each column represents a single regression. Significant and non-significant coefficients are designated at the top of the figure. Orange diamonds identify results from the true household coordinates using the bilinear extraction. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

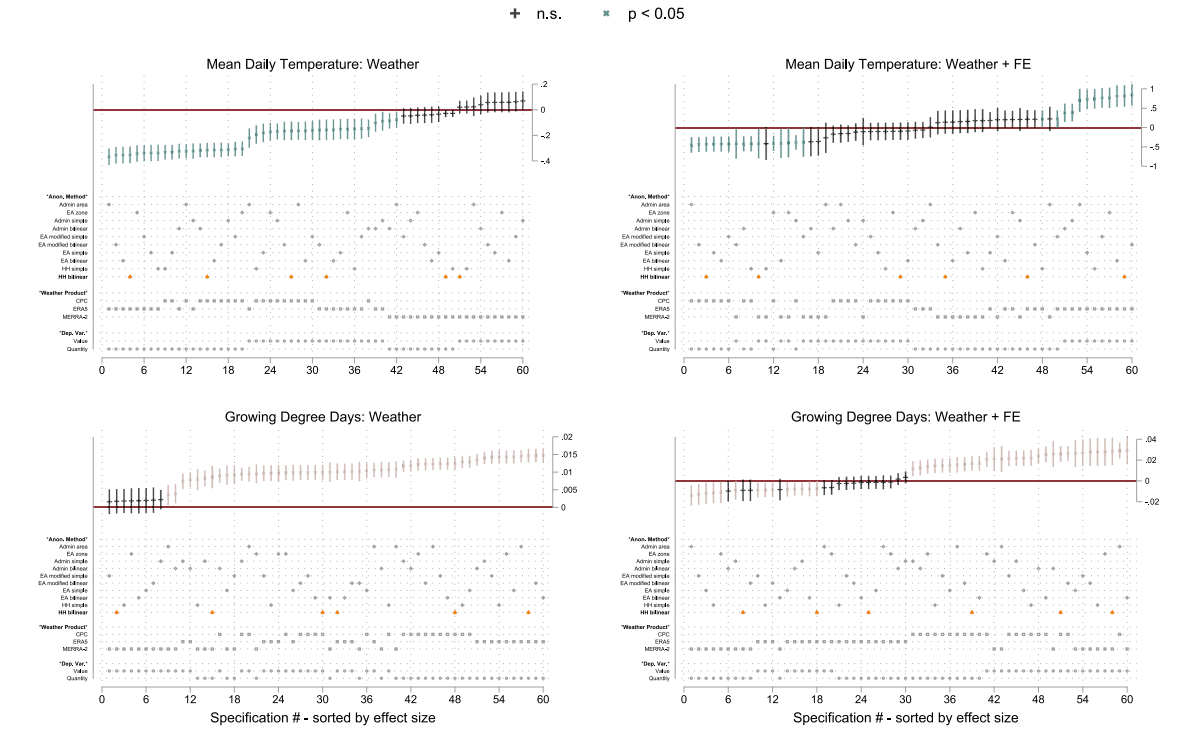

Fig. 20. Specification curve for temperature variables in Niger.
Note: The figure presents specification curves where each panel presents results from a different model. Each panel includes 60 regressions, where each column represents a single regression. Significant and non-significant coefficients are designated at the top of the figure. Orange diamonds identify results from the true household coordinates using the bilinear extraction. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

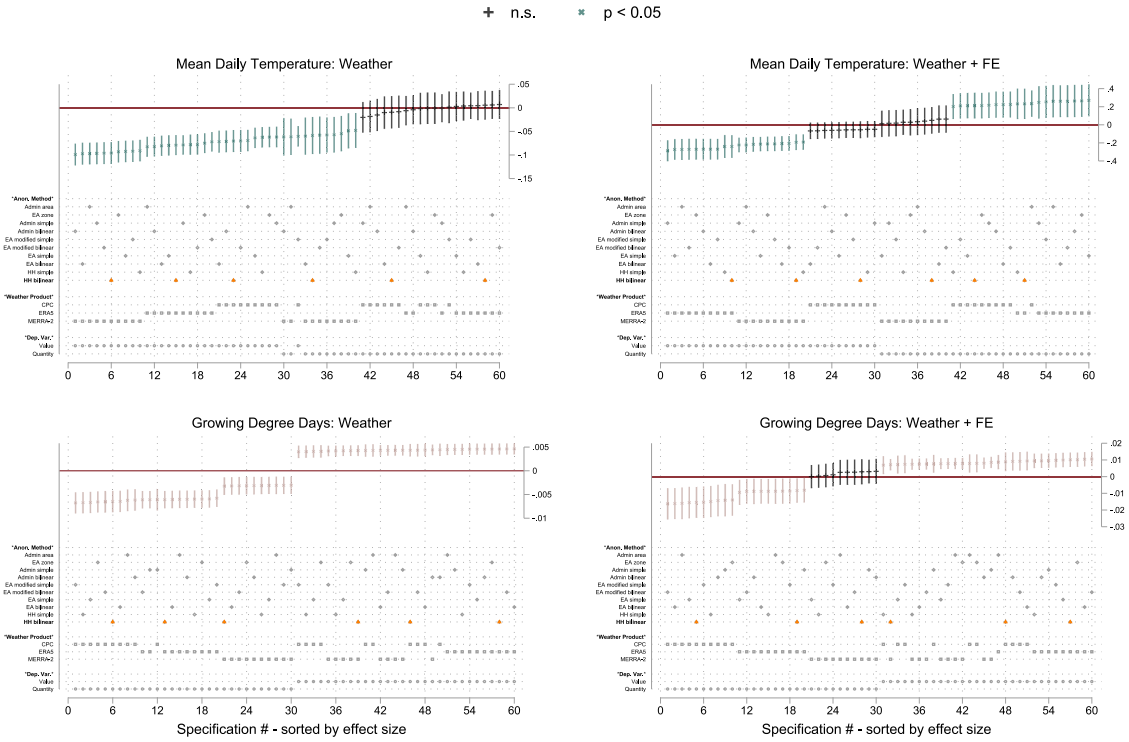

Fig. 21. Specification Curve for Temperature Variables in Nigeria.
Note: The figure presents specification curves where each panel presents results from a different model. Each panel includes 60 regressions, where each column represents a single regression. Significant and non-significant coefficients are designated at the top of the figure. Orange diamonds identify results from the true household coordinates using the bilinear extraction. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

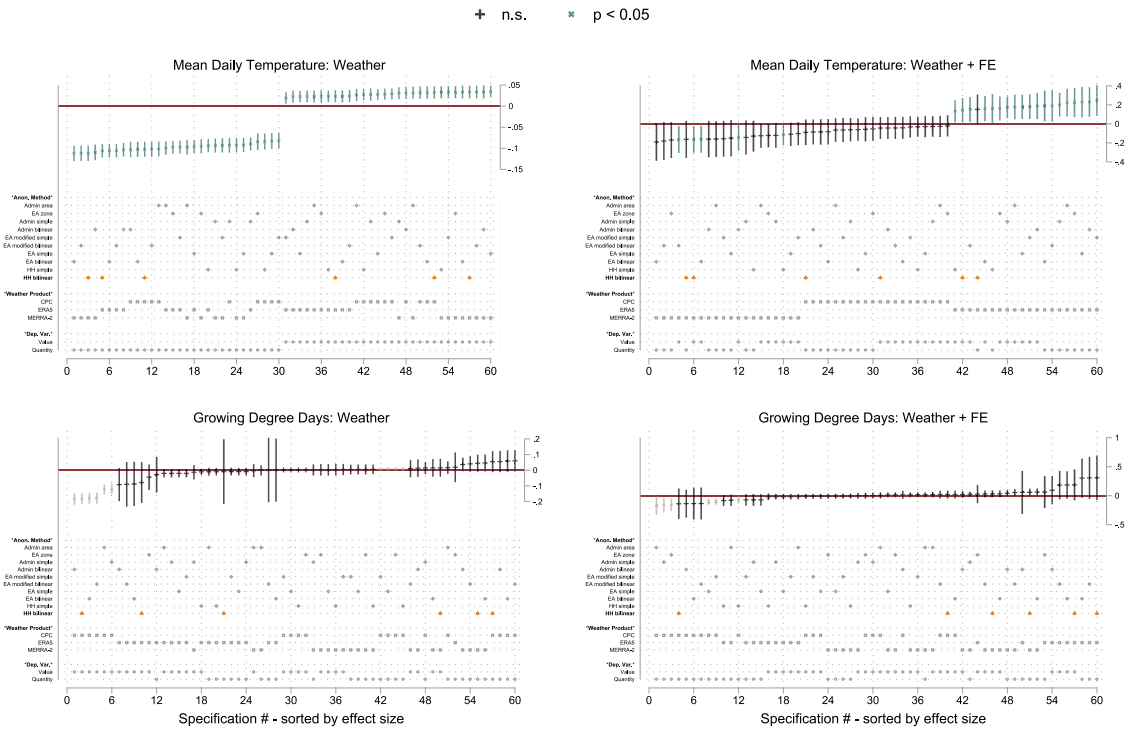

Fig. 22. Specification curve for temperature variables in Tanzania.
Note: The figure presents specification curves where each panel presents results from a different model. Each panel includes 60 regressions, where each column represents a single regression. Significant and non-significant coefficients are designated at the top of the figure. Orange diamonds identify results from the true household coordinates using the bilinear extraction. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

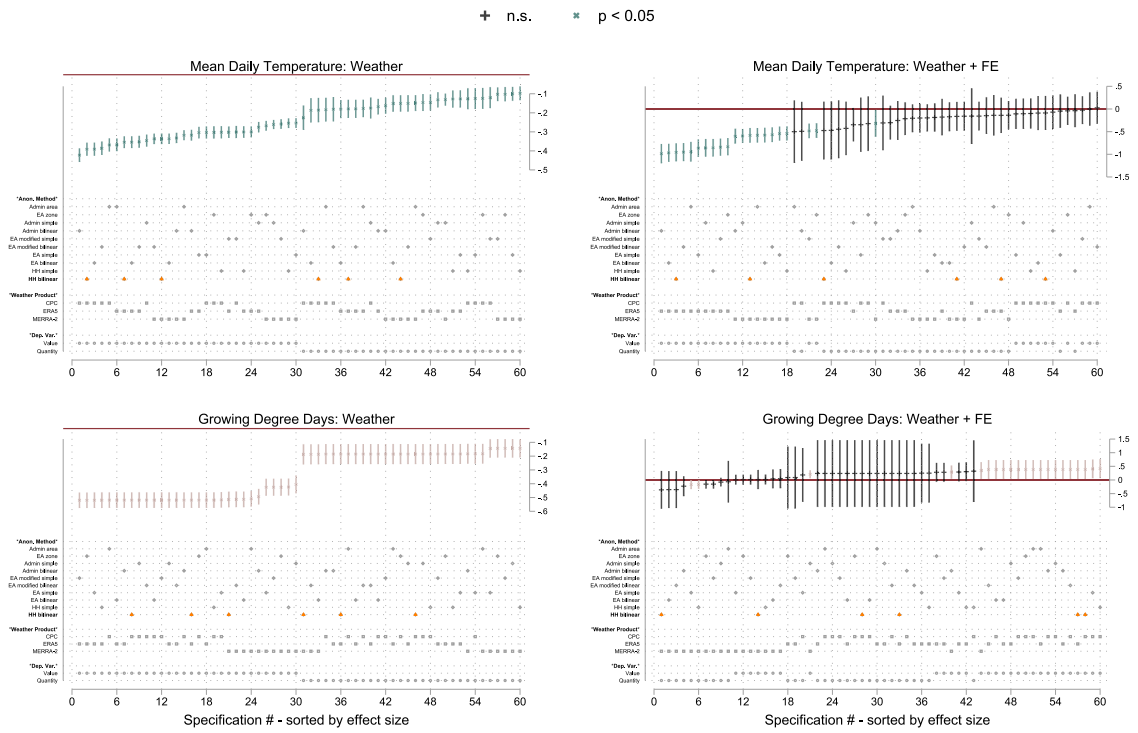

Fig. 23. Specification curve for temperature variables in Uganda.
Note: The figure presents specification curves where each panel presents results from a different model. Each panel includes 60 regressions, where each column represents a single regression. Significant and non-significant coefficients are designated at the top of the figure. Orange diamonds identify results from the true household coordinates using the bilinear extraction. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Ethiopia, Malawi, Niger, Nigeria, and Uganda. Only in Tanzania do results vary, with higher temperatures reducing yields but increasing the total value of harvest. For GDD, the metric is either positively correlated with outcomes or not significant in Malawi, Niger, and Uganda. In Ethiopia, Nigeria, and Tanzania, an increase in GDD can be either positively or negatively correlated with outcomes, depending on the remote sensing weather product that the data comes from and the dependent variable used in the regression.

When household and year fixed effects are added to the regressions, most temperature variables are no longer correlated with outcomes. The impact of including fixed effects varies by country and by temperature metric. As an example, in Ethiopia, without fixed effects, mean seasonal temperature is always negative or not significant but with fixed effects the correlation can be both positive (MERRA-2), negative (ERA5), or not significant (CPC). Conversely, GDD was both positively and negatively correlated with outcomes in Ethiopia, without fixed effects. Including fixed effects changes the results so that coefficients are always positively correlated or not significant. Similarly confounding patterns exist in Niger, Nigeria, Tanzania, and Uganda. Variables that were always of the same sign without fixed effects (mean and GDD in Niger and Uganda, mean in Nigeria) can have opposite signs when fixed effects are included. Or, variables that had opposite signs without fixed effects (mean and GDD in Tanzania) have consistent signs or are not significant when fixed effects are included. Which coefficients change signs with the inclusion of fixed effects is a function of both the source of the weather data and the choice of dependent variable. Only in Malawi do coefficients on temperature variables maintain consistent signs with and without fixed effects.

6. Towards a set of best practices

Having examined the results from 51,840 regressions on a panel survey database with 33,738 total household observations that span a decade and six countries in Eastern, Western, and Southern Africa with significant heterogeneity in agro-ecological conditions and rainfall patterns, it is useful to recapitulate the key takeaways towards the formulation of best practices and the identification of areas for future research.

Based on descriptive evidence and our heuristics, we find only minor evidence that SDL methods undertaken to protect privacy in the LSMS-ISA has an impact on the accuracy of results. The vast majority of spatial anonymization methods have no meaningful impact on estimates of the relationship between weather and agricultural productivity when compared to estimates from data that integrates weather and survey data using the exact household coordinates. To the extent that weak differences exist, they are in estimates from data that uses Administrative unit or Administrative area to match household locations to the gridded weather data products. Locations derived from administrative area provides the most privacy protection by introducing the most uncertainty regarding the exact location of a sampled household. And this privacy protect comes at a small cost in terms of data accuracy, resulting in some mismeasurement of the relationship between weather and agricultural productivity.

Though the results are generally robust to SDL methods to protect privacy, they are not robust to the choice of remote sensing weather product or the choice of weather metric. The correlation between rainfall or temperature and agricultural productivity varies by country depending on if the weather data comes from ARC2, CPC, CHIRPS, ERA5, MERRA-2, or TAMSAT. The relationship also varies depending on how one chooses to measure rainfall (e.g., mean daily or number of days without rain) and temperature (e.g., mean seasonal or GDD). Last, the relationship can vary depending on the choice of how to measure agricultural productivity (harvest value or yield). In extreme cases, the relationship between rainfall or temperature and agricultural productivity can have opposite signs depending on the source of the weather data, the metric to measure weather, and the metric to measure agricultural productivity. Although, we only briefly touch on these issues here, our populated pre-analysis plan explores these questions extensively (Michler et al., 2021a).

Remotely sensed weather data has become a common component of economic analysis (Dell et al., 2014; Donaldson and Storeygard, 2016). Yet, there has been little recognition in the economics literature that the need for privacy protection in public use survey data can introduce mismeasurement when integrating this data with remote sensing data. The need to protect privacy while producing accurate analysis has long been discussed in the computer science literature but has only recently been taken up in the economics literature (Abowd et al., 2019; Abraham, 2019; Chetty and Friedman, 2019; Ruggles et al., 2019). Neither has there been a convergence on a set of best practices for dealing with measurement error in the remote sensing data itself. Few empirical papers today would fail to verify the robustness of the results to different specifications (Simonsohn et al., 2020) or different iterations of the data (Steegen et al., 2016). Yet economics papers rarely, if ever, verify the robustness of results to the choice of remote sensing data source or weather metric.

In trying to formulate a set of best practices for researchers interested in the integration of public use survey data with publicly available remote sensing weather datasets we recommend the following:

1. At this time, researchers need not be concerned about potential inaccuracies that may be introduced into their analysis by integrating spatially anonymized survey datasets with publicly available remote sensing weather products. The current spatial resolution of the latter geospatial data is not fine enough for common SDL methods, such as geomasking, to result in mismeasurement of weather events that are experienced by sampled households.
2. Researchers must carefully choose which remote sensing source to use in their analysis. Despite the volume of precipitation and the temperature in a given location on a given day being objective facts, remote sensing products can differ substantially in how they measure these objective facts. Because of this, remote sensing products can and do disagree on what the weather was.
3. Researchers may want to demonstrate the robustness of their results to the choice of weather data drawn from different remote sensing products, or different weather metrics. When weather is critical to the identification strategy, results should not be sensitive to the choice of remote sensing product or the weather metric.

Despite the thematic focus of our paper on weather and agricultural productivity, future research should work towards building a robust body of knowledge regarding the impacts of using spatially anonymized survey data in a wide range of analytical and mapping applications. In specific cases, such as high-resolution crop area or crop yield mapping, it is clear that spatially anonymized public use datasets will not be useful, as researchers need access to survey data with precise agricultural plot locations for integration with higher-resolution satellite imagery, such as Sentinel-2 (Azzari et al., 2021). However, there is a high degree of thematic heterogeneity in research applications that rests on the integration of georeferenced socioeconomic survey datasets with geospatial data sources, and it is not always clear, ex-ante, to what extent, if any, spatial anonymization may lead to biased insights. A comprehensive body of evidence on the potential impacts of using spatially anonymized survey data will ultimately have implications for both survey data users and producers. While it can enable data users to better identify research questions whose answers may or may not be mediated by spatial anonymization of survey data, it can also provide further impetus for data producers to invest in physical and technological infrastructure to provide secure access to scientific use datasets that include confidential geolocation data that are not included in public use datasets but that may be needed to answer specific research questions.

CRediT authorship contribution statement

Jeffrey D. Michler: Conceptualization, Formal analysis, Methodology, Writing – original draft. **Anna Josephson:** Conceptualization, Formal analysis, Methodology, Writing – original draft. **Talip Kilic:** Conceptualization, Data curation, Funding acquisition, Writing – review & editing. **Siobhan Murray:** Conceptualization, Data curation, Writing – review & editing.

Data and code availability

All household and weather data used in this analysis are pubicly available. However, exact household GPS locations used to integrate household and weather data are private and held by the World Bank. Without the confidential household locations, the results in this paper cannot be replicated from raw data through cleaning to final analysis. To assist in replication, we have made the processed data and all code available at https://doi.org/10.5281/zenodo.6841500. This data and code allows for the reproduction of all results tables and figures in the published paper.

Appendix A. Supplementary data

Supplementary materials consisting of Appendices A-C referenced in the article can be found online at https://doi.org/10.1016/j.jdeveco.2022.102927.

References

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. J. Dev. Econ. 139, 171–184.

Abowd, J.M., Schmutte, I.M., 2015. Economic analysis and statistical disclosure limitation. Brook. Pap. Econ. Act. 46, 221–293.

Abowd, J.M., Schmutte, I.M., 2019. An economic analysis of privacy protection and statistical accuracy as social choice. Amer. Econ. Rev. 109 (1), 171–202.

Abowd, J.M., Schmutte, I.M., Sexton, W.N., Vilhuber, L., 2019. Why the economics profession must actively participate in the privacy protection debate. AEA Paers Proc. 109, 397–402.

Abraham, K.G., 2019. Reconciling data access and privacy: Building a sustainable model for the future. AEA Paers Proc. 109, 409–413.

Aragón, F.M., Oteiza, F., Pablo Rud, J., 2021. Climate change and agriculture: Subsistence farmers' response to extreme heat. Am. Econ. J: Econ. Policy 13 (1), 1–35.

Azzari, G., Jain, S., Jeffries, G., Kilic, T., Murray, S., 2021. Understanding the requirements for surveys to support satellite-based crop type mapping: Evidence from sub-Saharan Africa. Remote Sens. 13 (23), 4749.

Barrios, S., Bertinelli, L., Strobl, E., 2010. Trends in rainfall and economic growth in Africa: A neglected cause of the African growth tragedy. Rev. Econ. Stat. 92 (2), 350–366.

Blankespoor, B., Croft, T., Dontamsetti, T., Mayala, B., Murray, S., 2021. Spatial anonymization: Guidance note prepared for the inter-secretariat working group on household surveys. https://unstats.un.org/iswghs/task-forces/documents/Spatial_Anonymization_Report_submit01272021_ISWGHS.pdf.

Bosilovich, M., Lucchesi, R., Suarez, M., 2016. MERRA-2: File specification. GMAO Office Note No. 9 (Version 1.1), http://gmao.gsfc.nasa.gov/pubs/office_notes.

Brückner, M., Ciccone, A., 2011. Rain and the democratic window of opportunity. Econometrica 79 (3), 923–947.

Burke, M., Driscoll, A., Lobell, D.B., Ermon, S., 2021. Using satellite imagery to understand and promote sustainable development. Science 371 (6536), eabe8628.

Burke, M., Lobell, D.B., 2017. Satellite-based assessment of yield variation and its determinants in smallholder African systems. Proc. Natl. Acad. Sci. 114 (9), 2189–2194.

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

Central Statistics Agency of Ethiopia (CSA), 2014. Rural socioeconomic survey 2011–2012. Public Use Dataset. Ref: ETH_2011_ERSS_v01_M. Downloaded from https://microdata.worldbank.org/index.php/catalog/2053 on 6 September 2019.

Central Statistics Agency of Ethiopia (CSA), 2015. Ethiopia socioeconomic survey 2013–2014. Public Use Dataset. Ref: ETH_2013_ESS_v02_M. Downloaded from https://microdata.worldbank.org/index.php/catalog/2053 on 6 September 2019.

Central Statistics Agency of Ethiopia (CSA), 2017. Ethiopia socioeconomic survey, wave 3 (ESS3) 2015–2016. Public Use Dataset. Ref: ETH_2015_ESS_v02_M. Downloaded from https://microdata.worldbank.org/index.php/catalog/2783 on 6 September 2019.

Chen, J.J., Mueller, V., Jia, Y., Tseng, S.K.-H., 2017. Validating migration responses to flooding using satellite and vital registration data. Amer. Econ. Rev. 107 (5), 441–445.

Chen, M., Shi, W., Xie, P., Silva, V.B., Kousky, V.E., Higgins, R.W., Janowiak, J.E., 2008. Assessing objective techniques for gauge-based analyses of global daily precipitation. J. Geophys. Res. 113, D04110.

Chetty, R., Friedman, J.N., 2019. A practical method to reduce privacy loss when disclosing statistics based on small samples. AEA Paers Proc. 109, 414–420.

Corno, L., Hildebrandt, N., Voena, A., 2020. Age of marriage, weather shocks, and the direction of marriage payments. Econometrica 88 (3), 879–915.

Dell, M., Jones, B.F., Olken, B.A., 2012. Temperature shocks and economic growth: Evidence from the last half century. Am. Econ. J.: Macroecon. 4 (3), 66–95.

Dell, M., Jones, B.F., Olken, B.A., 2014. What do we learn from the weather? The new climate-economy literature. J. Econ. Lit. 52 (3), 740–798.

Deschêne, O., Greenstone, M., 2007. The economic impacts of climate change: Evidence from agricultural output and random fluctuations in weather. Amer. Econ. Rev. 97 (1), 354–385.

Dinur, I., Nissim, K., 2003. Revealing information while preserving privacy. In: Proceedings of ACM SIGMOD-SIGACT-Sigart Smyposium on Principles of Database Systems, Vol. 22. pp. 202–210.

Donaldson, D., Storeygard, A., 2016. The view from above: Applications of satellite data in economics. J. Econ. Perspect. 30 (4), 171–198.

Duflo, E., Banerjee, A., Finkelstein, A., Katz, L.F., Olken, B.A., Sautman, A., 2020. In praise of moderation: Suggestions for the scope and use of pre-analysis plans for RCTs in economics. NBER Working Paper 26993.

Dwork, C., McSherry, F., Nissim, K., Smith, A., 2006. Calibrating noise to sensitivity in private data analysis. In: Halevi, S., Rabin, T. (Eds.), Theory of Cryptography. Springer Berlin Heidelberg, pp. 265–284,

Dwyer-Lindgren, L., Cork, M.A., Sligar, A., Steuben, K.M., Wilson, K.F., Provost, N.R., Mayala, B.K., VanderHeide, J.D., Collison, M.L., Hall, J.B., Biehl, M.H., Carter, A., Frank, T., Douwes-Schultz, D., Burstien, R., Casey, D.C., Deshpande, A., Earl, L., Bcheraoui, C.E., .Farag, T.H., Henry, N.J., Kinyoki, D., Marczak, L.B., Nixon, M.R., Osgood-Zimmerman, A., Pigott, D., Jr., R.C.R., Ross, J.M., Schaeffer, L.E., Smith, D.L., Weaver, N.D., Wiens, K.E., Easton, J.W., Justman, J.E., Opio, A., Sartorius, B., Tanser, F., Wabiri, N., Piot, P., Murray, C.J., Hay, S.I., 2019. Mapping HIV prevalence in sub-Saharan Africa between 2000 and 2017. Nature 570, 189–193.

Funk, C., Peterson, P., Landsfeld, M., Pedreros, D., Verdin, J., Shukla, S., Husak, G., Rowland, J., Harrison, L., Hoell, A., Michaelsen, J., 2015. The climate hazards infrared precipitation with stations—A new environmental record for monitoring extremes. Sci. Data 2, 150066.

Garg, T., Jagnani, M., Taraz, V., 2020. Temperature and human capital in India. J. Assoc. Environ. Resourc. Econ. 7 (6), 1113–1150.

Gollin, D., Udry, C., 2021. Heterogeneity, measurement error, and misallocation: Evidence from African agriculture. J. Polit. Econ. 129 (1), 1–80.

Graetz, N., Friedman, J., Osgood-Zimmerman, A., Burstien, R., Biehl, M.H., Shields, C., Mosser, J.F., Casey, D.C., Deshpande, A., Earl, L., Reiner, R.C., Ray, S.E., Fullerman, N., Levine, A.J., Stubbs, R.W., Mayala, B.K., Longbottom, J., Browne, A.J., Bhatt, S., Weiss, D.J., Gething, P.W., Mokdad, A.H., Lim, S.S., Murray, C.J., Gakidou, E., Hay, S.I., 2018. Mapping local variation in education attainment across Africa. Nature 555, 48–53.

Hennermann, K., Berrisford, P., 2020. ERA5: data documentation. Last modified Nov 18, 2020. https://confluence.ecmwf.int/display/CKB/ERA5%3A+data+documentation.

Jagnani, M., Barrett, C.B., Liu, Y., You, L., 2021. Within-season producer response to warmer temperatures: Defensive investments by Kenyan farmers. Econ. J. 131 (633), 392–419.

Janzen, S.A., Michler, J.D., 2021. Ulysses' pact or Ulysses' raft: Using pre-analysis plans in experimental and nonexperimental research. Appl. Econ. Perspect. Policy forthcoming.

Jayachandran, S., 2006. Selling labor low: Wage responses to productivity shocks in developing countries. J. Polit. Econ. 114 (3), 538–575.

Jolliffe, D., Mahler, D.G., Veerappan, M., Kilic, T., Wollburg, P., 2021. Under what conditions are data valuable for development?. World Bank Policy Research Working Paper, No. 9811.

Kaur, S., 2019. Nominal wage rigidity in village labor markets. Amer. Econ. Rev. 109 (10), 3585–3616.

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 102, 102033.

König, M.D., Rohner, D., Thoenig, M., Zilibotti, F., 2017. Networks in conflict: Theory and evidence from the great war of Africa. Econometrica 85 (4), 1093–1132.

Kosmowski, F., Aragaw, A., Kilian, A., Ambel, A., Ilukor, J., Yigezu, B., Stevenson, J., 2019. Varietal identification in household surveys: Results from three household-based methods against the benchmark of DNA fingerprinting in southern Ethiopia. Exp. Agric. 55 (3), 371–385.

Levine, R., Renelt, D., 1992. A sensitivity analysis of cross-country growth regressions. Amer. Econ. Rev. 82 (4), 942–963.

Maccini, S., Yang, D., 2009. Under the weather: Health, schooling, and economic consequences of early-life rainfall. Amer. Econ. Rev. 99 (3), 1006–1026.

McCarthy, N., Kilic, T., Brubaker, J., Murray, S., 2021. Droughts and floods in Malawi: Impacts on crop production and the performance of sustainable land management practices under climate extremes. Environ. Dev. Econ. 26, 432–449.

Michler, J.D., Baylis, K., Arends-Kuenning, M., Mazvimavi, K., 2019a. Conservation agriculture and climate resilience. J. Environ. Econ. Manag. 93, 148–169.

Michler, J.D., Josephson, A., Kilic, T., Murray, S., 2019b. Empirically estimating the impact of weather on agriculture. http://dx.doi.org/10.17605/osf.io/z3snh, OSF Registries. July 1..

Michler, J.D., Josephson, A., Kilic, T., Murray, S., 2021a. Estimating the Impact of Weather on Agriculture. World Bank Policy Research Working Paper, No. 9867.

Michler, J.D., Josephson, A., Kilic, T., Murray, S., 2021b. Estimating the impact of weather on agriculture. arXiv.org. August 19 https://arxiv.org/abs/2012.11768v2.

Miguel, E., Satyanath, S., Sergenti, E., 2004. Economic shocks and civil conflict: An instrumental variables approach. J. Polit. Econ. 112 (4), 725–753.

Morten, M., 2019. Temporary migration and endogenous risk sharing in village India. J. Polit. Econ. 127 (1), 1–46.

National Bureau of Statistics (NBS), 2012. Nigeria General Household Survey (GHS), Panel 2010, wave 1. Public Use Dataset. Ref: NGA_2010_GHSP-W1_v04_M. Downloaded from https://microdata.worldbank.org/index.php/catalog/1002 on 6 September 2019.

National Bureau of Statistics (NBS), 2014. Nigeria General Household Survey, panel 2012–2013, wave 2. Public Use Dataset. Ref: NGA_2012_GHSP-W2_v02_M. Downloaded from https://microdata.worldbank.org/index.php/catalog/1952 on 6 September 2019.

National Bureau of Statistics (NBS), 2019. Nigeria General Household Survey, panel (GHS-panel) 2015–2016. Public Use Dataset. Ref: NGA_2015_GHSP-W3_v02_M. Downloaded from https://microdata.worldbank.org/index.php/catalog/2734 on 6 September 2019.

National Statistical Office (NSO), 2012. Third Malawi Integrated Household Survey 2010–2011. Public Use Dataset. Ref: MWI_2010_IHS-III_v01_M. Downloaded from https://microdata.worldbank.org/index.php/catalog/1003 on 6 September 2019.

National Statistical Office (NSO), 2015. Malawi Integrated Household Panel Survey 2010–2013 (short-term panel, 204 EAs). Public Use Dataset. Ref: MWI_2010-2013_IHPS_v01_M. Downloaded from https://microdata.worldbank.org/index.php/catalog/2248 on 6 September 2019.

National Statistical Office (NSO), 2017. Malawi Integrated Household Panel Survey 2010–2013-2016 (long-term panel, 102 EAs). Public Use Dataset. Ref: MWI_2010-2016_IHPS_v02_M. Downloaded from https://microdata.worldbank.org/index.php/catalog/2939 on 6 September 2019.

Novella, N.S., Thiaw, W.M., 2013. African rainfall climatology version 2 for famine early warning systems. J. Appl. Meteorol. Climatol. 52 (3), 588–606.

Osgood-Zimmerman, A., Millear, A.I., Stubbs, R.W., Shields, C., Pickering, B.V., Earl, L., Garaetz, N., Kinyoki, D.K., Ray, S.E., Bhatt, S., Browne, A.J., Burnstien, R., Cameron, E., .Casey, D.C., Despande, A., Fullman, N., Gething, P.W., Gibson, H.S., Henry, N.J., Herrero, M., Krause, L.K., Letourneau, I.D., Levine, A.J., Lui, P.Y., Longbottom, J., Mayala, B.K., Mosser, J.F., Noor, A.M., Pigott, D.M., Piwoz, E.G., Rao, P., Rawat, R., Reiner, R.C., Smith, D.L., Weiss, D.J., Wiens, K.E., Mokdad, A.H., Lim, S.S., Murray, C.J., Kassebaum, N.J., Hay, S.I., 2018. Mapping child growth failure in Africa between 2000 and 2015. Nature 555, 41–47.

Ruggles, S., Fitch, C., Magnuson, D., Schroeder, J., 2019. Differential privacy and census data: Implications for social and economic research. AEA Paers Proc 109, 403–408.

Sala-i-Martin, X.X., 1997a. I just ran four million regressions. NBER Working Paper 6252.

Sala-i-Martin, X.X., 1997b. I just ran two million regressions. Amer. Econ. Rev. 87 (2), 174–183.

Sarsons, H., 2015. Rainfall and conflict: A cautionary tale. J. Dev. Econ. 115, 62–72.

Shah, M., Steinberg, B.M., 2017. Drought of opportunities: Contemporaneous and long-term impacts of rainfall shocks on human capital. J. Polit. Econ. 125 (2), 527–561.

Simonsohn, U., Simmons, J.P., Nelson, L.D., 2020. Specification curve analysis descriptive and inferential statistics for all plausible specifications. Nat. Hum. Behav. 4, 1208–1214.

Steegen, S., Tuerlinckx, F., Gelman, A., Vanpaemel, W., 2016. Increasing transparency through a multiverse analysis. Perspect. Psychol. Sci. 11 (5), 702–712.

Suri, T., 2011. Selection and comparative advantage in technology adoption. Econometrica 79 (1), 159–209.

Survey and Census Division, National Institute of Statistics, Niger (NIS), 2014. National survey on household living conditions and agriculture 2011. Public Use Dataset. Ref: NER_2011_ECVMA_v01_M. Downloaded from https://microdata.worldbank.org/index.php/catalog/2050 on 6 September 2019.

Survey and Census Division, National Institute of Statistics, Niger (NIS), 2016. National survey on household living conditions and agriculture (ECVMA) 2014. Public Use Dataset. Ref: NER_2014_ECVMA_v02_M. Downloaded from https://microdata.worldbank.org/index.php/catalog/2676 on 6 September 2019.

Tanzania National Bureau of Statistics (TNBS), 2011. National Panel Survey 2008–2009, wave 1. Public Use Dataset. Ref: TZA_2008_NPS-R1_v03_M. Downloaded from https://microdata.worldbank.org/index.php/catalog/76 on 6 September 2019.

Tanzania National Bureau of Statistics (TNBS), 2012. National Panel Survey 2010–2011, wave 2. Public Use Dataset. Ref: TZA_2010_NPS-R2_v03_M. Downloaded from https://microdata.worldbank.org/index.php/catalog/1050 on 6 September 2019.

Tanzania National Bureau of Statistics (TNBS), 2015. National Panel Survey 2012–2013, wave 3. Public Use Dataset. Ref: TZA_2012_NPS-R3_v01_M. Downloaded from https://microdata.worldbank.org/index.php/catalog/2252 on 6 September 2019.

Taraz, V., 2018. Can farmers adapt to higher temperatures? Evidence from India. World Dev. 112, 205–219.

Tarnavsky, E., Grimes, D., Maidment, R., Black, E., Allan, R.P., Stringer, M., Chadwick, R., Kayitakire, F., 2014. Extension of the TAMSAT satellite-based rainfall monitoring over Africa and from 1983 to present. J. Appl. Meteorol. Climatol. 53 (12), 2805–2822.

Tesfaye, W., Blalock, G., Tirivayi, N., 2021. Climate-smart innovations and rural poverty in Ethiopia: Exploring impacts and pathways. Am. J. Agric. Econ. 103 (3), 878–899.

Uganda Bureau of Statistics (UBOS), 2014a. Uganda National Panel Survey (UNPS) 2010–2011. Public Use Dataset. Ref: UGA_2010_UNPS_v01_M. Downloaded from https://microdata.worldbank.org/index.php/catalog/2166 on 6 September 2019.

Uganda Bureau of Statistics (UBOS), 2014b. Uganda National Panel Survey (UNPS) 2010–2011. Public Use Dataset. Ref: UGA_2011_UNPS_v01_M. Downloaded from https://microdata.worldbank.org/index.php/catalog/2059 on 6 September 2019.

Uganda Bureau of Statistics (UBOS), 2019. National Panel Survey (UNPS) 2005–2009. Public Use Dataset. Ref: UGA_2005-2009_UNPS_v01_M. Downloaded from https://microdata.worldbank.org/index.php/catalog/1001 on 6 September 2019.

Wineman, A., Mason, N.M., Ochieng, J., Kirimi, L., 2017. Weather extremes and household welfare in rural Kenya. Food Secur. 9, 281–300.

Wood, A., Altman, M., Bembenek, A., Bun, M., Gaboardi, M., 2018. Differential privacy: A primer for a non-technical audience. Vanderbilt J. Entertain. Technol. Law 21 (1), 209–276.

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