Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Measuring what matters: Earth-observation measurement error, agricultural productivity and rural welfare in Sub-Saharan Africa

Oladimeji Abeeb Olaniyi

oladimeji_olaniyi@unilesa.edu.ng

University of Ilesa

Adewale Matthew Adekanmbi

Dominion University, Ibadan

Hammed Tiamiyu Olakulehin

Ladoke Akintola University of Technology

Research Article

Keywords: Earth observation, Measurement error, Agricultural productivity, Rural welfare, Common correlated effects, Statistical capacity, Sub-Saharan Africa

Posted Date: November 18th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11302884/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Measuring what matters: Earth-observation measurement error, agricultural productivity and rural welfare in Sub-Saharan Africa

ARTICLE INFO

Keywords: Earth observation; Measurement error; Agricultural productivity; Rural welfare; Common correlated effects; Statistical capacity; Sub-Saharan Africa

ABSTRACT

Satellite and reanalysis data are now the main source of weather information for studies of farming and poverty in Sub-Saharan Africa (SSA), yet different products often tell different stories about the same year. We ask how large this measurement error is, where it is largest, and how much it changes what we learn about the effect of weather on agricultural productivity and rural welfare. We build a panel of 48 SSA countries for 2001–2025 that links two widely used Earth-observation (EO) climate products, the ERA5 reanalysis and the station-based CRU TS dataset, to agricultural and welfare data from the World Bank, the IMF and the African Development Bank. The two products disagree on whether a year was wetter or drier than normal in more than four of every ten country-years, and this disagreement has grown over time. Using second-generation panel methods that allow for cross-sectional dependence and heterogeneous slopes, we show that rainfall and heat shocks shape crop and food production, value added per farm worker and cereal yields. However, estimates based on a single product understate these effects by between half and five-sixths once measurement error is corrected by using one product as an instrument for the other. Disagreement between products is larger in countries with weak statistical capacity and political instability, and attenuation is stronger there. Weather-driven changes in farm output do not show up clearly in national welfare indicators, which are themselves smoothed. We contribute to the literature on EO data in development economics by showing that measurement error is large, systematic and consequential, and we offer practical lessons for researchers and statistical agencies.

1. Introduction

Over the past two decades, research on agriculture and poverty in Sub-Saharan Africa (SSA) has increasingly shifted from ground-based statistics to data collected from space and generated by climate models. Satellite rainfall estimates, reanalysis products such as ERA5, and vegetation indices now cover the whole continent at fine resolution and at almost no cost to the user (Burke et al., 2021). As a result, Earth-observation (EO) data have become the most important means of measuring weather shocks in studies of farm output, food security and household welfare in low-income countries (Dell et al., 2014; Michler et al., 2022). This shift has been driven by improvements in satellite sensors and computing power, and by the weakness of ground observation networks in much of Africa, where many weather stations closed or stopped reporting after the 1980s (Harris et al., 2020).

In this way, EO data offer an important means for researchers and policy makers to study places where ground data are thin (Burke et al., 2021). Indeed, there is a sense that satellites have solved the problem of missing data, and that a researcher's ability to study African agriculture increasingly depends on remote sensing rather than on national statistics (Jean et al., 2016; Lobell et al., 2020). There is also the hopeful view that EO data are objective and free from the political and administrative problems that affect official figures (Henderson et al., 2012; Jerven, 2013). Indeed, the advice to researchers has been to replace self-reported or official weather and output data with satellite measures wherever possible.

While recent literature shows that EO data can predict crop yields and poverty with reasonable accuracy, little is known about how much the choice of EO product changes the answers to basic economic questions. Different products combine satellite signals, ground stations and model physics in different ways, so they can disagree about the same place and year (Auffhammer et al., 2013). Michler et al. (2022) found that the estimated effect of weather on farm output in household surveys changes with the rainfall and temperature product used. Understanding this problem is critical because measurement error in an explanatory variable bias its estimated effect towards zero, so that weather shocks may look less harmful than they really are (Bound et al., 2001; Hausman, 2001).

However, for EO measurement error to be taken seriously, researchers need to know how large it is, where it is concentrated and whether it can be corrected. Measurement error is unlikely to be the same everywhere. Where ground stations are sparse, satellite and reanalysis products have little local information to anchor them, and their errors are likely to be larger (Harris et al., 2020; Hersbach et al., 2020). Because we are interested in understanding measurement error as a problem of data systems, we focus on the link between EO disagreement and national statistical capacity, which reflects the wider ability of a country to collect and publish reliable data (Dang et al., 2023).

While the move to EO data may appeal to researchers working on countries with poor statistics, evidence is lacking on whether it removes measurement problems or simply moves them. Official agricultural statistics in Africa are known to be weak and are often revised (Carletto, Jolliffe, & Banerjee, 2015; Jerven, 2013). Welfare indicators such as undernourishment and child mortality are often modelled estimates that change slowly from year to year (Deaton, 2005). Therefore, researchers may adopt EO data in the hope of better measurement but instead combine noisy weather data with noisy outcome data, and draw weak or misleading conclusions.

In addition, the development literature explains that statistical systems depend on institutions and resources (Devarajan, 2013), but relatively weak statistical capacity is common in SSA. The measurement literature explains that errors in both explanatory and outcome variables can be addressed if a second, independent measure of the same variable is available (Griliches & Hausman, 1986; Schennach, 2016). Therefore, we ask: How large is measurement error in EO weather data for SSA, how does it affect estimates of the effect of weather on agricultural productivity and rural welfare, and is it larger where statistical capacity is weak?

To explore this problem, we ground our research in the literature on weather and economic outcomes (Dell et al., 2014; Hsiang, 2016), paying particular attention to those studies that explicitly considered measurement error in EO data (e.g., Auffhammer et al., 2013; Michler et al., 2022) and the role of data quality in African agricultural statistics (e.g., Carletto, Jolliffe, & Banerjee, 2015; Gollin & Udry, 2021). This follows the call from Michler et al. (2022) for a closer analysis of how the choice of EO product affects economic estimates, and from Burke et al. (2021) for more work on the limits of satellite data in development research.

Our study uses a panel of 48 SSA countries over 2001–2025. It combines two of the most widely used EO climate products, the ERA5 reanalysis (Hersbach et al., 2020) and the station-based CRU TS dataset (Harris et al., 2020), both taken from the World Bank Climate Change Knowledge Portal, with agricultural and welfare data from the World Bank World Development Indicators (WDI), the IMF and the African Development Bank (AfDB). The dataset allows us to measure how much the two products disagree, to correct for this disagreement using one product as an instrument for the other, and to relate the size of the error to national statistical capacity.

This study makes three key contributions to the development economics literature. First, it advances the literature on EO data by revealing the size and growth of measurement error in weather data for SSA. While EO data offer wide coverage and consistency, the two products we study disagree on the sign of the annual rainfall anomaly in more than four of every ten country-years, and their agreement has weakened over time. This insight invites a more careful treatment of EO data as measurements with error, not as ground truth.

Second, the study shows how much this error matters for economic estimates. Using second-generation panel estimators that allow for cross-sectional dependence and heterogeneous slopes, we show that single-product estimates of the effect of rainfall on crop production, food production and farm value added are between one-sixth and one-half of the error-corrected estimates, and that the effect of heat on cereal yields is almost invisible in one product but large and significant after correction.

Third, we contribute to the literature on statistical capacity by showing that EO measurement error is not random across countries. Disagreement between products is larger, and attenuation is stronger, where statistical capacity and political stability are weak. This suggests that satellites do not remove the data gap between countries but partly reproduce it.

2. Literature review

2.1. Earth-observation data

EO data are defined as information about the earth's surface and atmosphere that is gathered by satellites, aircraft and ground sensors and processed into consistent measures over space and time (Burke et al., 2021). In economics, the most common EO products are satellite-based rainfall estimates, gridded station datasets and climate reanalyses (Auffhammer et al., 2013). In subsequent work, researchers refined the concept by stressing that most "satellite" weather data are in fact models: they combine satellite signals with ground observations and physical equations to produce a best estimate of conditions at each point (Hersbach et al., 2020). As such, EO data allow us to analyse weather at both a "global" scale, with consistent methods everywhere, and a "local" scale, with fine spatial detail.

EO products have multiplied and evolved rapidly over the past decades. Gridded station datasets such as CRU TS interpolate records from weather stations onto a regular grid, so their quality depends on how many stations report in each area (Harris et al., 2020). Reanalyses such as ERA5 feed satellite, radiosonde and station observations into a weather forecasting model to reconstruct past conditions every hour (Hersbach et al., 2020). Satellite rainfall products such as CHIRPS blend infrared satellite data with station records (Funk et al., 2015). Each approach has advantages: station data are close to the ground where they exist, reanalyses are physically consistent, and satellites see everywhere.

The composition of these products is carefully designed to fill gaps in ground data, but this also means that their errors differ. Where stations are dense, gridded station data are accurate; where they are sparse, the grid is filled by interpolation from distant stations or by long-term averages (Harris et al., 2020). Reanalyses are less dependent on local stations, but they inherit biases from their model physics, especially for rainfall in the tropics (Hersbach et al., 2020). While EO products improve the coverage of weather data, they come at the cost of hidden measurement error, which differs across products, places and years (Auffhammer et al., 2013).

2.2. Sub-Saharan Africa and Earth-observation measurement

From an SSA perspective, the ability to measure what matters for rural livelihoods depends on data that are often weak or missing. Researchers (e.g., Devarajan, 2013; Jerven, 2013) argue that African official statistics suffer from low funding, old survey frames and political pressure, a situation described as "Africa's statistical tragedy". The measurement concept is central to empirical economics, where it reflects the gap between the variable that theory calls for and the variable that can be observed. As such, measurement error is defined as the difference between the measured value of a variable and its true value (Bound et al., 2001).

From an agricultural perspective, measurement problems appear in both inputs and outputs. Land area, harvest quantities and yields are often estimated by farmer recall or by extension officers, and these estimates contain large errors (Carletto, Gourlay, & Winters, 2015; Desiere & Jolliffe, 2018). Errors in land and output measurement can even create false relationships, such as the inverse relationship between farm size and productivity (Abay et al., 2019). Gollin and Udry (2021) show that much of the apparent dispersion in productivity across African farms is due to measurement error and unobserved heterogeneity rather than to misallocation.

Independent of the source, measurement error matters because it changes estimated effects. Classical measurement error in an explanatory variable, which is random and unrelated to the true value, biases its coefficient towards zero, a problem known as attenuation (Bound et al., 2001; Hausman, 2001). Non-classical error, which is correlated with the true value or with other variables, can bias estimates in either direction (Abay et al., 2019; Schennach, 2016). Furthermore, errors in the dependent variable reduce precision, and if they are related to the explanatory variable, they also bias estimates (Hausman, 2001). As a result, measurement error not only makes estimates less precise but may also change the conclusions drawn from them.

EO data were expected to reduce these problems. Satellite measures of weather are not affected by farmer recall or by the incentives of statistical agencies, and they can be used to predict yields and poverty (Jean et al., 2016; Lobell et al., 2020). Interestingly, studies have shown that EO data contain their own errors, which are sometimes as large as those in survey data. Lobell et al. (2020) found that satellite yield estimates in Uganda performed as well as farmer reports only when calibrated with ground data. Michler et al. (2022) found that estimated weather effects in household surveys varied substantially across EO products and with the spatial masking used to protect privacy.

The importance of data quality is supported by scholars who study statistical capacity. The World Bank Statistical Capacity Indicator and its successor, the Statistical Performance Indicators, measure the ability of a country to produce timely and reliable statistics (Dang et al., 2023). Countries with weak statistical capacity tend to have fewer functioning weather stations, fewer agricultural surveys and more frequent revisions to official data (Carletto, Jolliffe, & Banerjee, 2015). Thus, measurement error in EO data is not "mechanical" but can be influenced by the conditions of the national data system.

Looking at the consequences for policy, recent literature clarifies that measurement error can hide real effects. If rainfall or heat is measured with error, the estimated damage from weather shocks will be understated, and the case for adaptation investment will look weaker than it should (Auffhammer et al., 2013; Dell et al., 2014). However, the literature is mostly silent on how large this bias is for SSA as a whole, how it has changed over time, and whether it is larger in some countries than in others. By comparing two independent EO products, one can measure the error directly and correct for it.

2.3. Weather shocks, agricultural productivity and rural welfare

In this study, we operationalise weather shocks as departures of annual rainfall and temperature from each country's own 1981–2010 climate. Capturing departures is important because farmers adapt to their normal climate, so it is unusual weather that causes shocks to output and income (Dell et al., 2014; Hsiang, 2016). At the same time, such shocks are hard to measure accurately where ground data are poor, making the theoretical exploration of this relationship particularly salient (e.g., Michler et al., 2022).

We theorise that four mechanisms link weather shocks to agricultural productivity and rural welfare in SSA. First, rainfall determines soil moisture and crop growth in rain-fed systems, which account for most cropland in the region (Schlenker & Lobell, 2010). Years with below-normal rainfall reduce yields, planted area and livestock output, while years with near-normal or above-normal rainfall raise them (Barrios et al., 2010). Some very wet years can also cause floods and crop losses, so the relationship may not be linear.

Second, high temperatures damage crops directly by shortening the growing season, raising water stress and reducing grain filling (Schlenker & Roberts, 2009). Evidence from Africa suggests that warming reduces the yields of maize, sorghum and millet, with large losses expected under future climate change (Lobell et al., 2011; Schlenker & Lobell, 2010). While farmers can adapt by changing planting dates and varieties, such adaptation is costly and incomplete, particularly for poor farmers with little access to credit or information.

Third, shocks to farm output reduce rural incomes and food availability. Because most rural households in SSA depend on their own production for food and income, a poor harvest reduces consumption and raises food insecurity (Dercon, 2004; Hill & Porter, 2017). Households try to cope by selling assets, drawing on savings or migrating, but these strategies are often insufficient in severe or repeated shocks (Dercon, 2004). In contrast, good harvests can raise incomes and lower food prices (Sen, 1981).

Fourth, the effect of weather on welfare may be slower and more diffuse than its effect on output. National welfare indicators such as undernourishment and child mortality depend on many factors, including health services, prices and imports, and many are modelled estimates that smooth year-to-year changes (Deaton, 2005). Given that weather shocks mainly work through farm output, we expect their effects on welfare to appear through agricultural productivity rather than directly. These arguments lead us to separate productivity and welfare outcomes. Therefore, we posit:

H1a. *Adverse weather shocks measured with Earth-observation data reduce agricultural productivity in Sub-Saharan Africa.*

H1b. *Adverse weather shocks measured with Earth-observation data reduce rural welfare in Sub-Saharan Africa.*

Fig. 1 depicts our theoretical model.

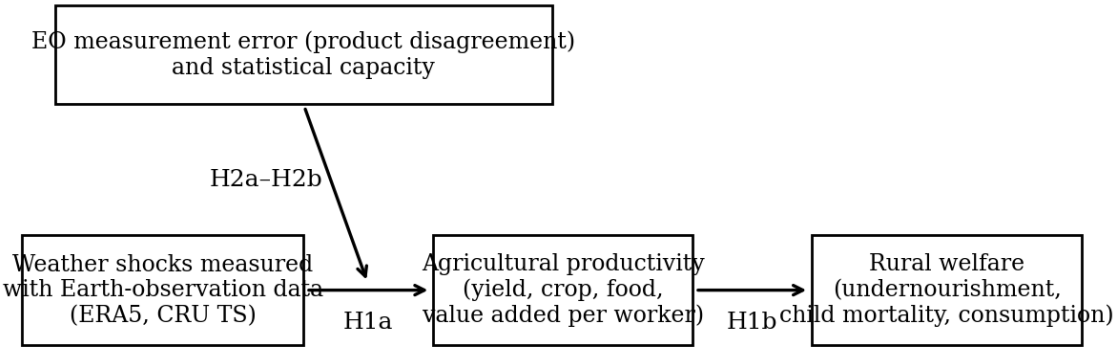


Fig. 1. Empirical model.

2.4. Measurement error and statistical capacity

Measurement is central to theorising about weather and development because a researcher's conclusions depend on the ability to observe both the shock and its consequences (Bound et al., 2001). The econometric literature shows that when a variable is measured with error, its effect is understated, and the understatement grows with the share of the measured variance that is noise rather than signal (Griliches & Hausman, 1986; Hausman, 2001). Thus, measurement error potentially links data quality to the size of estimated effects (Schennach, 2016).

Two independent measures of the same variable provide an important means to correct this bias. If the errors in two measures are unrelated to each other, one can be used as an instrument for the other, which removes the attenuation caused by classical error (Griliches & Hausman, 1986). ERA5 and CRU TS are built from different inputs and methods: ERA5 relies on a weather model fed by satellite and other observations, while CRU TS interpolates station records (Harris et al., 2020; Hersbach et al., 2020). Their errors are therefore unlikely to be strongly related, which makes them suitable instruments for each other, although they may share some ground stations.

While EO data are seen as a way around weak statistics, they may also reflect the weakness of national data systems. Gridded station products such as CRU TS depend directly on national weather services for station data, and reanalyses depend on the same observations to anchor their models (Harris et al., 2020; Hersbach et al., 2020). Where statistical and meteorological services are weak, fewer observations enter these products, and the products fall back on interpolation, long-term averages or model physics. In SSA, where station networks have declined, this may produce larger disagreement between products in countries with weak statistical capacity (Dang et al., 2023; Jerven, 2013).

However, as we previously explained, we argue that the consequences of this disagreement for economic research remain underexplored. Researchers usually choose one product, often for convenience, and rarely report how their results would change with another (Auffhammer et al., 2013; Michler et al., 2022). If measurement error is larger in countries with weak statistical capacity, estimates of weather effects will be most attenuated precisely in the countries where policy needs them most. Scholars examining measurement in African agriculture argue that errors are systematic rather than random, and that they are related to the capacity of the data system (Carletto, Jolliffe, & Banerjee, 2015; Gollin & Udry, 2021).

Taking the problem of measurement error and the role of statistical capacity into account, we argue that EO measurement error weakens the estimated relationship between weather shocks and agricultural productivity, and that this weakening is stronger where statistical capacity is low, and posit our second hypothesis:

H2a. *Earth-observation measurement error attenuates the estimated relationship between weather shocks and agricultural productivity in Sub-Saharan Africa.*

H2b. *The attenuation caused by Earth-observation measurement error is stronger in countries with weak statistical capacity.*

3. Methods

While EO data are acknowledged as a key resource for studying agriculture and welfare in SSA (Burke et al., 2021), the measurement error they contain has not been satisfactorily quantified for the region as a whole (Michler et al., 2022). This is partly due to the limited number of studies that compare products across many countries and years, exacerbated by the difficulty of obtaining consistent outcome data and by the fact that measurement errors change over time, thus requiring longitudinal studies.

3.1. Sample

We tested our hypotheses using annual panel data for the 48 SSA countries in the World Bank's regional classification over 2001–2025. The data were collected in September 2026 from recognised international databases. Climate data come from the World Bank Climate Change Knowledge Portal, which provides country averages of the ERA5 0.25° reanalysis (Hersbach et al., 2020) and of the CRU TS 0.5° gridded station dataset, versions 4.09 and 4.08 (Harris et al., 2020). Agricultural, welfare and statistical capacity data come from the World Bank WDI and its archived vintages, governance data from the Worldwide Governance Indicators (WGI) (Kaufmann et al., 2011), macroeconomic data from the IMF World Economic Outlook, and infrastructure data from the AfDB Socio-Economic Database.

Because CRU TS ends in 2024 and the FAO production indices reported in WDI end in 2022, the panel is unbalanced. In total, the main estimation sample comprises between 1,034 and 1,068 country-year observations for 44 to 48 countries, depending on the outcome. The number of observations by subregion summarised in Tables 1 and the number of observations by statistical capacity group is summarized in Table 2. The number of observations in each period is shown in Table 3.

Table 1
Number of country-year observations by subregion.

Subregion	Countries	2001–2005	2006–2010	2011–2015	2016–2020	2021–2025	Total
Central Africa	9	45	45	45	45	18	198
East Africa	14	60	60	68	70	28	286
Southern Africa	9	45	45	45	45	18	198
West Africa	16	80	80	80	80	32	352
Total	48	230	230	238	240	96	1034

Note: Observations with the crop production index and both Earth-observation climate products available. CRU TS ends in 2024 and the FAO production indices in 2022, which explains the smaller 2021–2025 count. Source: Authors' calculation

Table 2
Number of country-year observations by statistical capacity group.

Statistical capacity group	Countries	2001–2005	2006–2010	2011–2015	2016–2020	2021–2025	Total
Low statistical capacity	23	110	110	114	115	46	495
High statistical capacity	24	120	120	120	120	48	528
Not rated	1	0	0	4	5	2	11

Note: Groups split at the median of the World Bank Statistical Capacity Indicator averaged over 2004–2010. Source: Authors' calculation

Table 3
Number of observations in each period.

Period	Number of countries	Number of observations
2001–2005	46	230
2006–2010	46	230
2011–2015	48	238
2016–2020	48	240
2021–2025	48	96
Total observations		1034

Source: Authors' calculation

3.2. Measures

3.2.1. Dependent variable

The performance-dependent variables in our study are agricultural productivity and rural welfare. Agricultural productivity is measured in four ways, all in logs: cereal yield in kilograms per hectare, the crop production index, agriculture value added per worker in constant 2015 dollars, and the food production index (Barrios et al., 2010; Lobell et al., 2011). These measures capture land productivity, total crop output, labour productivity and total food output, respectively. Rural welfare is measured by the prevalence of undernourishment, under-five mortality (in logs) and household final consumption per capita (in logs), which capture food deprivation, health and living standards (Deaton, 2005; Hill & Porter, 2017). To measure errors in official statistics, we also compute the absolute log difference between cereal yields reported in the current WDI and in the April 2016 WDI vintage for the same country and year.

3.2.2. Independent variables

Weather shocks measured with EO data are our independent variables. For each product, the rainfall anomaly is annual rainfall minus its 1981–2010 country mean, divided by its 1981–2010 standard deviation, and the temperature anomaly is the annual mean temperature minus its 1981–2010 country mean, in °C (Dell et al., 2014; Hsiang, 2016). Each product is standardised against its own climatology, so that a systematic difference in levels between products does not count as disagreement. The international literature usually uses one product at a time (Barrios et al., 2010; Dell et al., 2012). However, to understand measurement error as a concept, we use both ERA5 and CRU TS and measure their disagreement as the absolute difference between their anomalies. We also measure the revision of CRU TS rainfall between versions 4.08 and 4.09 as a measure of vintage error in the same product. The values of the independent variables are measured in the same year as the outcome, since weather affects the current harvest.

Our moderating variable, statistical capacity, is the World Bank Statistical Capacity Indicator (SCI, 0–100) averaged over 2004–2010, before most of the outcomes we study. Countries below the median are classified as having low statistical capacity. We also use the Statistical Performance Indicators overall score (SPI, 0–100), available from 2016, and the WGI index of political stability as alternative measures of the conditions under which data are produced (Dang et al., 2023; Kaufmann et al., 2011).

3.2.3. Control variables

Because weather anomalies are determined by nature and not by economic choices, they are as good as randomly assigned once country effects and common shocks are removed (Dell et al., 2014; Hsiang, 2016). Adding economic controls such as fertiliser use, prices or income would therefore introduce "bad controls" that are themselves affected by weather. Instead, we control for unobserved common factors, such as world food prices, regional climate patterns and global recessions, by including the cross-sectional averages of the outcome and of the weather variables, with a separate coefficient for each country (Pesaran, 2006). These averages absorb shocks that affect all countries but to different degrees, and they also absorb country-specific trends driven by common factors (Kapetanios et al., 2011). The variables used in the analysis are summarised in Table 4.

Table 4
Explanation of the variables used in the analysis.

Variables	References	Explanation
Dependent variables		
Cereal yield	Lobell et al. (2011)	Log of kilograms per hectare of harvested land (WDI, FAO)
Crop production index	Barrios et al. (2010)	Log of crop production index, 2014–2016 = 100 (WDI, FAO)
Agriculture value added per worker	Gollin and Udry (2021)	Log of constant 2015 US$ (WDI)
Food production index	Hill and Porter (2017)	Log of food production index, 2014–2016 = 100 (WDI, FAO)
Undernourishment	Hill and Porter (2017)	Prevalence of undernourishment, % of population (WDI, FAO)
Under-five mortality	Deaton (2005)	Log of deaths per 1,000 live births (WDI, UN IGME)
Household consumption	Deaton (2005)	Log of household final consumption per capita, constant 2015 US$ (WDI)
Independent variables		
Rainfall anomaly (ERA5 / CRU TS)	Dell et al. (2014); Hsiang (2016)	Annual rainfall minus 1981–2010 mean, divided by 1981–2010 SD, each product (CCKP)
Temperature anomaly (ERA5 / CRU TS)	Schlenker and Roberts (2009)	Annual mean temperature minus 1981–2010 mean, °C, each product (CCKP)
Rainfall disagreement	Michler et al. (2022)	Absolute difference between ERA5 and CRU TS rainfall anomalies (SD)
Temperature disagreement	Auffhammer et al. (2013)	Absolute difference between ERA5 and CRU TS temperature anomalies (°C)
CRU vintage revision	Harris et al. (2020)	Absolute % difference in rainfall between CRU TS 4.09 and 4.08
Moderating and other variables		
Statistical capacity	Dang et al. (2023)	World Bank Statistical Capacity Indicator, 0–100, mean 2004–2010; low = below median (WDI archives)
Statistical performance	Dang et al. (2023)	Statistical Performance Indicators overall score, 0–100 (WDI)
Political stability	Kaufmann et al. (2011)	WGI political stability and absence of violence, −2.5 to 2.5
Official data revision	Carletto, Jolliffe and Banerjee (2015)	Absolute log difference (×100) between current and April 2016 WDI cereal yield

Source: Authors' calculation

3.3. Statistical model

In our sample, observations are nested within countries and years, and countries are linked by common shocks; thus, the statistical models needed to account for both heterogeneity across countries and dependence between them. Therefore, we applied second-generation panel estimators, starting with the common correlated effects model of Pesaran (2006):

$$Y_{it} = \alpha_i + \beta' W_{it} + \delta_i' \bar{Z}_t + \varepsilon_{it} \qquad (1)$$

where i denotes the country, t denotes the year, α_i is the country-specific effect, W is the vector of weather anomalies (rainfall and temperature), $\bar{Z}_t$ is the vector of cross-sectional averages of the outcome and of the weather variables, δ_i are country-specific loadings, and ε is the error term. We estimate equation (1) with the pooled CCE estimator (CCEP), which pools the weather coefficients but allows each country to respond differently to the common factors, and with the CCE mean group estimator (CCEMG), which estimates a separate β_i for each country and averages them.

To correct for EO measurement error (H2a), we use the errors-in-variables approach of Griliches and Hausman (1986): the ERA5 anomalies are instrumented with the CRU TS anomalies within the CCEP framework,

$$W^{ERA5}_{it} = \pi_i + \Pi\, W^{CRU}_{it} + \lambda_i' \bar{Z}_t + v_{it} \qquad (2)$$

which removes the attenuation bias caused by classical errors that are independent across products. The ratio of the single-product coefficient to the corrected coefficient gives the attenuation ratio, which equals the share of the measured variance that reflects true weather (Bound et al., 2001). To test whether attenuation is stronger where statistical capacity is weak (H2b), we add interactions between the weather anomalies and a low-capacity indicator, instrument them with the corresponding CRU TS interactions, and compare attenuation ratios across subsamples.

Using cross-sectional averages allowed us to control for unobserved common factors that affect countries to different degrees, such as the 2008 food price crisis, El Niño events and the COVID-19 recession (Pesaran, 2006). In addition, allowing for country-specific loadings addresses the heterogeneity of SSA economies and remains valid when the common factors are non-stationary (Kapetanios et al., 2011). For robustness, we also estimate the augmented mean group estimator (AMG) of Eberhardt and Teal (2011), the cross-sectionally augmented autoregressive distributed lag model (CS-ARDL) of Chudik and Pesaran (2015) and Chudik et al. (2016), which gives long-run effects, and first-generation two-way fixed effects with Driscoll and Kraay (1998) standard errors for comparison. All estimations were performed in Python with the statsmodels and linearmodels libraries.

3.3.1. Bias testing

We first checked the second-generation properties of the data. The Pesaran (2015) test shows strong cross-sectional dependence in every variable (Table A1 in the Appendix), which confirms that countries are linked by common shocks and that first-generation estimators would give misleading standard errors. After the CCE transformation, the residual CD statistic falls from between 19.4 and 74.0 to between −1.9 and 3.5, and it is insignificant at the 5 percent level for the crop, food and value added models, showing that the cross-sectional averages capture most of the common factors. The slope homogeneity test of Pesaran and Yamagata (2008) rejects equal slopes across countries for every outcome and product (Δ adjusted between 3.15 and 8.13, $p < 0.01$), which supports the use of estimators that allow heterogeneity.

Next, we tested for unit roots with the CIPS test of Pesaran (2007), which allows for cross-sectional dependence. The four productivity measures and all weather and disagreement variables are stationary, while undernourishment, under-five mortality and household consumption are non-stationary in levels. CCE estimators remain consistent in the presence of non-stationary common factors (Kapetanios et al., 2011), but we treat the welfare results with extra caution and also report the long-run CS-ARDL estimates.

Another concern was the validity of the instruments used to correct measurement error. Because ERA5 and CRU TS are built from different inputs, their errors are unlikely to be strongly related, but they may share some station data. The first-stage F statistics range from 58 to 134 for the productivity models, far

above the conventional threshold of 10 (Stock & Yogo, 2005), so weak instruments are not a problem. If the errors of the two products are positively related, our corrected estimates are a lower bound on the true effect, since shared errors would move the instrumental variables estimate back towards the attenuated one (Bound et al., 2001). For the welfare models, in which official crop production is instrumented with weather from both products, the Hansen J test does not reject the overidentifying restrictions (p between 0.205 and 0.469).

4. Results

4.1. Descriptive analysis

The two EO products often disagree about the weather (Table 5a). Using the pooled terciles of the absolute difference between the ERA5 and CRU TS rainfall anomalies, the share of country-years with high disagreement (more than 1.22 standard deviations) rose from 17.9 percent in 2001–2005 to 50.5 percent in 2021–2024. Table 5b shows why. ERA5 records rainfall below its 1981–2010 normal in every period (on average −0.33 standard deviations), while CRU TS records rainfall at or above normal (on average +0.34 standard deviations). The products report opposite signs for the rainfall anomaly in 35.4 percent of country-years in 2001–2005 and in 52.6 percent in 2021–2024, and the within-country correlation of their anomalies fell from 0.63 to 0.41. The products also differ on warming: by 2021–2024, ERA5 records an average temperature anomaly of 0.77 °C, compared with 0.41 °C in CRU TS. Even the same product changes between versions: CRU TS rainfall for the same country and year differs by about 1 percent on average between versions 4.08 and 4.09.

Table 5a
Rainfall disagreement between Earth-observation products across the sampled period.

Level of disagreement	2001–2005	2006–2010	2011–2015	2016–2020	2021–2025	Total
Low (≤ 0.51 SD)	110 (45.8)	71 (29.6)	81 (33.8)	76 (31.7)	46 (24.0)	384 (33.3)
Medium (0.51–1.22 SD)	87 (36.2)	91 (37.9)	80 (33.3)	77 (32.1)	49 (25.5)	384 (33.3)
High (> 1.22 SD)	43 (17.9)	78 (32.5)	79 (32.9)	87 (36.2)	97 (50.5)	384 (33.3)
Total	240 (100)	240 (100)	240 (100)	240 (100)	192 (100)	1152 (100)

Note: Percentage in parenthesis. Disagreement is the absolute difference between the ERA5 and CRU TS rainfall anomalies (standard deviations of each product's 1981–2010 climate); cut-offs are the pooled terciles. 2021–2025 covers 2021–2024 because CRU TS ends in 2024. Source: Authors' calculation

Table 5b
Earth-observation climate products across the sampled period (country averages).

EO indicator	2001–2005	2006–2010	2011–2015	2016–2020	2021–2025
Rainfall anomaly, ERA5 (SD)	-0.48	-0.22	-0.41	-0.00	-0.51
Rainfall anomaly, CRU TS (SD)	0.02	0.45	0.15	0.54	0.57
Temperature anomaly, ERA5 (°C)	0.22	0.28	0.34	0.67	0.77
Temperature anomaly, CRU TS (°C)	0.25	0.28	0.23	0.38	0.41
Rainfall disagreement, abs(ERA5 − CRU) (SD)	0.73	1.00	0.95	1.09	1.48
Temperature disagreement, abs(ERA5 − CRU) (°C)	0.16	0.17	0.24	0.36	0.44
Years with opposite rainfall signs (%)	35.42	44.58	41.25	47.08	52.60
Within-country correlation of rainfall anomalies	0.63	0.56	0.57	0.43	0.41
CRU vintage revision, TS 4.08 to 4.09 (%)	0.80	0.84	1.56	1.01	0.71

Source: Authors' calculation from the World Bank Climate Change Knowledge Portal (ERA5 0.25° reanalysis; CRU TS 4.09 and 4.08, 0.5°).

Regarding the outcomes (Table 6), agricultural productivity rose over the period. Cereal yields increased from about 1,250 to 1,610 kilograms per hectare, the crop production index from 73.9 to 119.6, and value added per agricultural worker from about 1,820 to 2,610 dollars. Welfare indicators improved as well: undernourishment fell from 24.3 to 19.9 percent and under-five mortality from 117 to 61 deaths per 1,000 live births. However, official data are also measured with error: cereal yields for the same country and year differ between the current WDI and the 2016 vintage by 2.8 percent on average for 2001–2005 and 11.7 percent for 2011–2015.

Table 6
Agricultural productivity, rural welfare and official data revisions across the sampled period (country averages).

Indicator	2001–2005	2006–2010	2011–2015	2016–2020	2021–2025
Cereal yield (kg per hectare)	1,249	1,396	1,395	1,451	1,608
Crop production index (2014–2016 = 100)	73.92	83.40	96.73	108.73	119.62
Agriculture value added per worker (constant 2015 US$)	1,818	2,013	2,226	2,478	2,607
Food production index (2014–2016 = 100)	73.68	84.68	97.41	106.90	116.96
Prevalence of undernourishment (%)	24.30	20.90	19.64	19.44	19.93
Under-five mortality (per 1,000 live births)	117.21	94.83	78.27	68.50	61.40
Household consumption per capita (constant 2015 US$)	1,089	1,308	1,424	1,577	1,630
Revision to official cereal yield since WDI 2016 (%)	2.82	3.69	11.73	–	–

Note: The revision is the absolute log difference (×100) between the cereal yield reported in the current WDI and in the April 2016 WDI vintage for the same country and year. Source: Authors' calculation from WDI (FAO data) and WDI archives.

Table 7 presents the correlations between our variables. The Spearman correlation between the ERA5 and CRU TS rainfall anomalies is only 0.339, and that between their temperature anomalies is 0.525, which shows that the two products share less than half of their variation. Rainfall anomalies are positively correlated with the crop production index in both products (0.142 and 0.153). Rainfall disagreement is negatively correlated with baseline statistical capacity (−0.139) and political stability (−0.223), giving a first sign that measurement error is larger where data systems and institutions are weaker.

Table 7
Descriptive statistics and correlation matrix.

	Variable	Mean	S.D.	1	2	3	4	5	6	7	8	9	10	11	12
1	Cereal yield (log)	7.023	0.760												
2	Crop production index (log)	4.502	0.282	0.199***											
3	Ag. value added per worker (log)	7.214	0.988	0.086**	0.115***										
4	Undernourishment	20.903	12.457	-0.276***	-0.128***	-0.529***									
5	Under-five mortality (log)	4.288	0.597	-0.236***	-0.428***	-0.465***	0.318***								
6	Household consumption (log)	6.832	0.853	0.147***	0.110**	0.814***	-0.466***	-0.600***							
7	Rainfall anomaly, ERA5	-0.325	1.101	0.095**	0.142***	0.101***	-0.043	-0.202***	0.157***						
8	Rainfall anomaly, CRU TS	0.336	0.926	-0.024	0.153***	-0.039	0.044	0.015	-0.158***	0.339***					
9	Temperature anomaly, ERA5	0.453	0.400	-0.007	0.296***	-0.013	-0.065*	-0.199***	-0.028	-0.249***	-0.009				
10	Temperature anomaly, CRU TS	0.305	0.270	0.080**	0.091**	0.146***	-0.220***	-0.152***	0.092**	-0.067*	-0.135***	0.525***			
11	Rainfall disagreement	1.033	0.837	-0.055	0.055	-0.107***	0.045	0.071*	-0.202***	-0.345***	0.357***	0.198***	0.023		
12	Statistical capacity (baseline)	56.779	13.334	0.299***	0.020	-0.009	-0.292***	-0.266***	0.028	0.122***	-0.026	-0.050	0.102***	-0.139***	
13	Political stability	-0.633	0.893	0.114***	-0.035	0.338***	-0.284***	-0.472***	0.517***	0.271***	-0.153***	-0.204***	0.022	-0.223***	0.310***

*Note: Spearman correlation; *, **, *** denote p<0.05, p<0.01, and p<0.001. Source: Authors' calculation*

4.2. Regression analysis

We began by testing the main hypothesis that weather shocks affect agricultural productivity (H1a) and that EO measurement error attenuates this effect (H2a). Table 8 shows that a one standard deviation increase in rainfall raises the crop production index by 2.0 percent with either product, but by 3.7 percent once measurement error is corrected ($p = 0.010$). For the food production index, the effect rises from 1.0 percent with ERA5 ($p = 0.099$) and 1.6 percent with CRU TS to 3.0 percent after correction ($p = 0.009$). For value added per agricultural worker, the ERA5 estimate is close to zero (0.4 percent, $p = 0.563$), the CRU TS estimate is 1.6 percent, and the corrected estimate is 2.8 percent ($p = 0.032$). The attenuation ratios for ERA5 are 0.52, 0.33 and 0.16, respectively, meaning that the single-product estimates capture between one-sixth and one-half of the corrected effect.

The results for cereal yields show the role of temperature. With ERA5, the temperature effect on yields is zero (−0.002, $p = 0.987$); with CRU TS, a 1 °C warmer year lowers yields by 11.2 percent (coefficient −0.119, $p = 0.012$); after correction, the loss is 22.6 percent (coefficient −0.256, $p < 0.001$). The rainfall effect on yields, in contrast, is imprecise in all three columns. Therefore, H1a is supported: rainfall raises crop output, food output and labour productivity in agriculture, and heat lowers cereal yields. H2a is also supported: single-product estimates are strongly attenuated. The first-stage F statistics (58 to 134) confirm that the corrections are not driven by weak instruments.

Table 8
Common correlated effects regression on the relationship between Earth-observation weather shocks and agricultural productivity.

Panel A	(1a) ERA5 Coefficient (p-value)	(1b) CRU TS Coefficient (p-value)	(1c) Error-corrected Coefficient (CI 95%)	p-value	(2a) ERA5 Coefficient (p-value)	(2b) CRU TS Coefficient (p-value)	(2c) Error-corrected Coefficient (CI 95%)	p-value
Outcome	Cereal yield (log)				Crop production index (log)			
Explanatory variables								
Rainfall anomaly (SD)	0.065 (0.156)	0.019 (0.075)	0.034 (-0.015, 0.082)	0.176	0.020 (0.006)	0.020 (0.002)	0.037 (0.009, 0.066)	0.010
Temperature anomaly (°C)	-0.002 (0.987)	-0.119 (0.012)	-0.256 (-0.384, -0.129)	<0.001	-0.025 (0.231)	-0.037 (0.108)	-0.038 (-0.099, 0.022)	0.212
Diagnostics								
Attenuation ratio, rainfall (ERA5 ÷ corrected)	1.94				0.52			
First-stage F (rainfall / temperature)			58.4 / 69.6				73.4 / 133.8	
Cross-sectional averages (country-specific loadings)	Yes	Yes	Yes		Yes	Yes	Yes	
Observations	1068	1068	1068		1034	1034	1034	
Countries	46	46	46		48	48	48	

Panel B	(3a) ERA5 Coefficient (p-value)	(3b) CRU TS Coefficient (p-value)	(3c) Error-corrected Coefficient (CI 95%)	p-value	(4a) ERA5 Coefficient (p-value)	(4b) CRU TS Coefficient (p-value)	(4c) Error-corrected Coefficient (CI 95%)	p-value
Outcome	Ag. value added per worker (log)				Food production index (log)			
Explanatory variables								
Rainfall anomaly (SD)	0.004 (0.563)	0.016 (0.009)	0.028 (0.002, 0.053)	0.032	0.010 (0.099)	0.016 (0.002)	0.030 (0.008, 0.052)	0.009
Temperature anomaly (°C)	-0.016 (0.600)	0.008 (0.731)	0.028 (-0.050, 0.105)	0.485	-0.018 (0.303)	-0.033 (0.065)	-0.031 (-0.081, 0.019)	0.230
Diagnostics								
Attenuation ratio, rainfall (ERA5 ÷ corrected)	0.16				0.33			
First-stage F (rainfall / temperature)			91.2 / 110.6				72.6 / 128.7	
Cross-sectional averages (country-specific loadings)	Yes	Yes	Yes		Yes	Yes	Yes	
Observations	1060	1060	1060		1034	1034	1034	
Countries	44	44	44		48	48	48	

Notes: Pooled common correlated effects (CCEP) estimator with country-specific loadings on the cross-sectional averages of the outcome and regressors (Pesaran, 2006). Columns (a) and (b) use ERA5 and CRU TS weather, respectively; columns (c) are errors-in-variables estimates in which the ERA5 measures are instrumented with the CRU TS measures. Standard errors clustered by country. Source: Authors' calculation

We then estimated the effect of weather shocks on rural welfare (H1b). The results (Table 9, Panel A) show no significant direct effects of rainfall or temperature on undernourishment, under-five mortality or household consumption, either with ERA5 or after correction. Panel B shows that official crop production is linked to welfare: a 10 percent higher crop production index is associated with about 0.5 percentage points lower undernourishment (coefficient −5.145, $p < 0.001$). However, when crop production is instrumented with EO weather, so that only weather-driven changes in output are used, the effect is no longer significant, and the first-stage F statistics (14.0 to 18.8) are weaker than in the productivity models. Therefore, H1b is not supported at the national annual level. As we discuss below, this is likely to reflect measurement problems in the welfare indicators themselves, many of which are modelled and smoothed, as much as a true absence of effects.

Table 9
Common correlated effects regression on the relationship between Earth-observation weather shocks, agricultural productivity and rural welfare.

	(5) Undernourishment (%)		**(6) Under-five mortality (log)**		**(7) Household consumption per capita (log)**	
	Coefficient (CI 95%)	p-value	Coefficient (CI 95%)	p-value	Coefficient (CI 95%)	p-value
Panel A: Weather shocks (reduced form)						
Rainfall anomaly, ERA5	-0.121 (-0.357, 0.114)	0.312	0.001 (-0.006, 0.009)	0.709	-0.001 (-0.009, 0.006)	0.710
Temperature anomaly, ERA5	-0.584 (-1.415, 0.247)	0.168	0.016 (-0.019, 0.052)	0.370	0.016 (-0.011, 0.043)	0.237
Rainfall anomaly, error-corrected	0.138 (-0.560, 0.837)	0.698	0.003 (-0.015, 0.020)	0.772	0.002 (-0.021, 0.025)	0.873
Temperature anomaly, error-corrected	-1.758 (-4.460, 0.944)	0.202	-0.008 (-0.095, 0.079)	0.859	0.043 (-0.062, 0.149)	0.418
Panel B: Crop production (official statistics)						
Crop production index (log), CCEP	-5.145 (-8.145, -2.144)	<0.001	0.027 (-0.036, 0.091)	0.399	0.046 (-0.040, 0.132)	0.297
Crop production index (log), EO-instrumented	1.756 (-7.214, 10.726)	0.701	-0.063 (-0.431, 0.304)	0.736	-0.224 (-0.598, 0.150)	0.240
First-stage F	18.8		16.7		14.0	
Hansen J test, p-value	0.205		0.469		0.404	
Observations (Panel A / Panel B)	966 / 924		1152 / 1034		952 / 821	

Notes: CCEP estimator with country-specific loadings on cross-sectional averages; standard errors clustered by country. In Panel B, the crop production index is instrumented with the ERA5 and CRU TS rainfall and temperature anomalies. Source: Authors' calculation

Finally, we examined whether attenuation is stronger where statistical capacity is weak (H2b). The results in Table 10 show that, in the single-product (CCEP) estimates, the interactions between rainfall and low statistical capacity are negative for all three outcomes (−0.018 for crop production, −0.006 for value added per worker and −0.021 for food production, $p = 0.053$), which would suggest that low-capacity countries are less sensitive to rainfall. Once measurement error is corrected, however, these interactions shrink towards zero (0.002, 0.026 and −0.014, none significant). This indicates that much of the apparent difference between country groups reflects measurement error rather than a true difference in sensitivity. The subsample attenuation ratios point in the same direction: for crop production, single-product estimates capture 37 percent of the corrected effect in low-capacity countries but 75 percent in high-capacity countries; for value added per worker and food production, the ratios are close to zero in low-capacity countries but 0.47 and 0.57 in high-capacity countries.

The lower panel of Table 10 shows that disagreement between products is related to the quality of data systems. A 10-point higher Statistical Capacity Indicator is associated with 0.08 standard deviations less rainfall disagreement ($p = 0.051$), and a 10-point higher Statistical Performance Index with 0.18 standard deviations less rainfall disagreement ($p = 0.042$) and 0.10 °C less temperature disagreement ($p = 0.018$). A one-point higher political stability score is associated with 0.24 standard deviations less rainfall disagreement ($p < 0.001$). Thus, H2b is partially supported: measurement error is systematically larger where statistical capacity and stability are weaker, and attenuation is stronger there, although the interaction terms are not individually significant after correction.

Table 10
Moderation impact of statistical capacity on the relationship between Earth-observation weather shocks and agricultural productivity.

	(8) Crop production index (log)			(9) Ag. value added per worker (log)			(10) Food production index (log)		
	CCEP Coefficient (p-value)	Error-corrected Coefficient (CI 95%)	p-value	CCEP Coefficient (p-value)	Error-corrected Coefficient (CI 95%)	p-value	CCEP Coefficient (p-value)	Error-corrected Coefficient (CI 95%)	p-value
Explanatory variables									
Rainfall anomaly	0.028 (0.005)	0.036 (-0.006, 0.078)	0.097	0.009 (0.255)	0.020 (-0.008, 0.049)	0.157	0.020 (0.012)	0.034 (0.002, 0.066)	0.036
Temperature anomaly	-0.009 (0.661)	-0.011 (-0.087, 0.064)	0.770	-0.002 (0.949)	-0.002 (-0.082, 0.078)	0.968	-0.004 (0.829)	-0.006 (-0.070, 0.059)	0.864
Interaction terms									
Rainfall anomaly × low statistical capacity	-0.018 (0.185)	0.002 (-0.052, 0.056)	0.946	-0.006 (0.668)	0.026 (-0.031, 0.082)	0.372	-0.021 (0.053)	-0.014 (-0.055, 0.027)	0.515
Temperature anomaly × low statistical capacity	-0.038 (0.446)	-0.084 (-0.224, 0.056)	0.239	-0.033 (0.597)	0.064 (-0.117, 0.245)	0.485	-0.033 (0.409)	-0.073 (-0.172, 0.025)	0.144
Subsample attenuation ratio, rainfall									
Low statistical capacity	0.37			0.01			-0.15		
High statistical capacity	0.75			0.47			0.57		
Observations	1023			1060			1023		

Drivers of EO disagreement	Rainfall disagreement Coefficient (CI 95%)	p-value	Observations	Temperature disagreement Coefficient (CI 95%)	p-value	Observations
Statistical Capacity Indicator (0–100)	-0.0078 (-0.0156, 0.0000)	0.051	802	-0.0018 (-0.0044, 0.0007)	0.160	802
Statistical Performance Index (0–100)	-0.0180 (-0.0353, -0.0007)	0.042	402	-0.0103 (-0.0187, -0.0018)	0.018	402
Political stability (WGI)	-0.2416 (-0.3701, -0.1131)	<0.001	1095	-0.0280 (-0.0785, 0.0225)	0.277	1095

Notes: Upper panel: CCEP estimator; low statistical capacity = 1 for countries below the median baseline Statistical Capacity Indicator (2004–2010). The attenuation ratio is the ERA5 coefficient divided by the error-corrected coefficient in each subsample; values below one indicate attenuation. Lower panel: panel regressions with year effects, one regressor at a time, standard errors clustered by country. Source: Authors' calculation

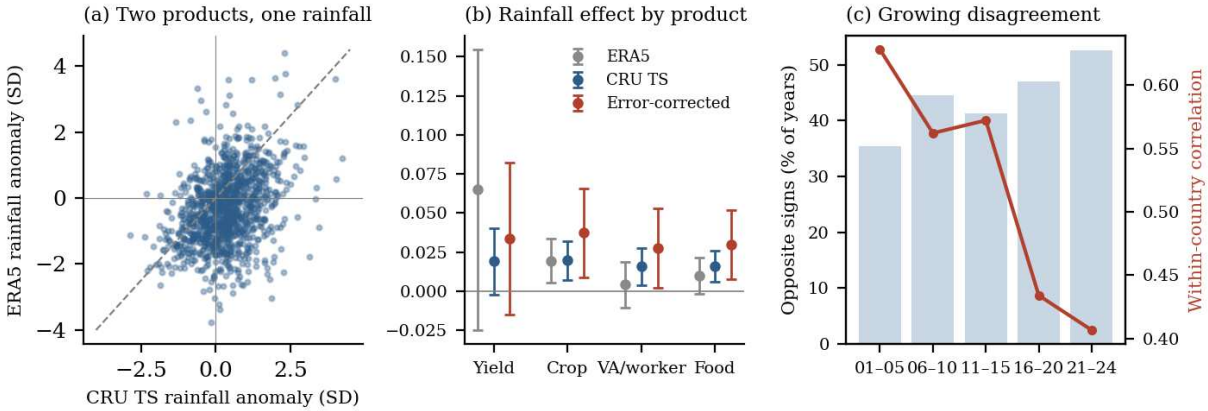


Fig. 2. Graphical display of Earth-observation measurement error: (a) ERA5 against CRU TS rainfall anomalies; (b) estimated rainfall effects on agricultural productivity by product and after error correction (95 % confidence intervals); (c) share of country-years with opposite rainfall signs and within-country correlation of anomalies by period.

4.3. Robustness check

We tested the robustness of our main results with alternative second-generation estimators (Table 11). The CCE mean group estimates, which allow each country its own slope, are close in size to the pooled estimates but less precise, as expected with heterogeneous slopes and about 20 years per country. The AMG estimator gives significant rainfall effects on crop production (1.9 percent with ERA5 and 1.7 percent with CRU TS) and cereal yields, and significant temperature effects on yields (−9.7 percent and −12.3 percent). The CS-ARDL long-run estimates have the same signs, and the error-correction coefficients (−0.39 to −0.59, $p < 0.001$) show that productivity returns to its long-run path within about two years after a shock.

The first-generation two-way fixed effects estimates reveal how much product choice can matter when cross-sectional dependence is ignored. For cereal yields, the temperature effect is positive with ERA5 (0.078) but negative and significant with CRU TS (−0.098); for value added per worker, it is negative with ERA5 (−0.028) but positive and significant with CRU TS (0.163). The same research question could therefore produce opposite answers depending on the product chosen. Averaging the two products gives estimates between the two single-product results, as expected when averaging reduces but does not remove noise. Using the rainfall deficit instead of the anomaly, instrumenting ERA5 with the older CRU TS 4.08 vintage, and excluding small island states all leave the main conclusions unchanged: rainfall matters for crop output, food output and farm labour productivity, heat matters for yields, and error-corrected effects are larger than single-product effects. These tests lead us to conclude that our results are robust.

Table 11
Robustness checks with alternative second-generation estimators and measures.

Estimator / specification	Cereal yield (log): Rainfall	Temperature	Crop production index (log): Rainfall	Temperature	Ag. value added per worker (log): Rainfall	Temperature	Food production index (log): Rainfall	Temperature
CCE mean group (CCEMG), ERA5	0.046 [0.025]	-0.073 [0.047]	0.013 [0.009]	-0.039 [0.029]	0.015 [0.010]	-0.064 [0.044]	0.006 [0.007]	-0.029 [0.021]
CCE mean group (CCEMG), CRU TS	0.028 [0.017]	-0.086 [0.051]	0.013 [0.008]	-0.076* [0.036]	0.026* [0.013]	-0.003 [0.046]	0.015* [0.006]	-0.064* [0.029]
Augmented mean group (AMG), ERA5	0.033** [0.012]	-0.097** [0.034]	0.019*** [0.005]	-0.029 [0.016]	0.002 [0.008]	-0.029 [0.018]	0.011* [0.005]	-0.007 [0.010]
Augmented mean group (AMG), CRU TS	0.028* [0.011]	-0.123* [0.053]	0.017** [0.006]	-0.049** [0.017]	0.018 [0.012]	-0.015 [0.022]	0.014*** [0.004]	-0.029** [0.011]
CS-ARDL long-run, ERA5	0.056 [0.042]	-0.162 [0.160]	0.048 [0.037]	-0.087 [0.055]	0.020 [0.014]	-0.017 [0.093]	0.027 [0.093]	-0.103 [0.068]
CS-ARDL long-run, CRU TS	-0.011 [0.039]	-0.151 [0.149]	0.019 [0.040]	-0.078 [0.052]	0.035 [0.021]	-0.027 [0.085]	0.024* [0.012]	-0.077* [0.034]
Two-way FE, Driscoll–Kraay (first generation), ERA5	0.057** [0.020]	0.078 [0.067]	0.016** [0.005]	0.035 [0.023]	0.015* [0.006]	-0.028 [0.021]	0.011 [0.006]	0.040 [0.028]
Two-way FE, Driscoll–Kraay (first generation), CRU TS	0.026** [0.008]	-0.098* [0.044]	0.030*** [0.006]	-0.114*** [0.029]	0.022** [0.007]	0.163** [0.051]	0.025*** [0.006]	-0.096** [0.032]
Average of ERA5 and CRU TS (CCEP)	0.058 [0.037]	-0.059 [0.081]	0.027** [0.008]	-0.028 [0.024]	0.015* [0.007]	0.003 [0.030]	0.016** [0.006]	-0.024 [0.019]
Rainfall deficit, ERA5 (CCEP)	-0.082 [0.051]	-0.005 [0.090]	-0.029* [0.011]	-0.028 [0.022]	-0.010 [0.012]	-0.017 [0.030]	-0.016 [0.009]	-0.020 [0.018]
Rainfall deficit, error-corrected	-0.072 [0.039]	-0.238*** [0.059]	-0.040 [0.027]	-0.038 [0.037]	-0.005 [0.019]	-0.003 [0.037]	-0.030 [0.022]	-0.033 [0.028]
Error-corrected with CRU TS 4.08 vintage as instrument	0.061 [0.041]	-0.236*** [0.063]	0.038* [0.015]	-0.038 [0.031]	0.035* [0.017]	0.040 [0.039]	0.030** [0.012]	-0.031 [0.025]
Excluding small island states, ERA5 (CCEP)	0.017 [0.010]	-0.086* [0.037]	0.019* [0.009]	-0.031 [0.025]	0.000 [0.010]	-0.014 [0.038]	0.008 [0.007]	-0.026 [0.020]
Excluding small island states, error-corrected	0.030 [0.027]	-0.129* [0.057]	0.038* [0.017]	-0.056 [0.034]	0.032* [0.015]	0.057 [0.055]	0.030* [0.013]	-0.036 [0.027]
CS-ARDL error-correction speed, ERA5	-0.593*** [0.050]		-0.557*** [0.046]		-0.389*** [0.035]		-0.540*** [0.048]	

*Notes: *, **, *** denote p<0.05, p<0.01, p<0.001; standard errors in brackets. CCEMG and AMG standard errors are the non-parametric mean-group standard errors; CS-ARDL(1,0) long-run coefficients are trimmed mean-group estimates (5 %); rainfall-deficit rows report the deficit coefficient (a negative value means a drier year lowers productivity). Source: Authors' calculation*

5. Discussion

5.1. Earth-observation weather shocks and agricultural productivity

While the development literature reports that weather shocks reduce farm output in SSA (Barrios et al., 2010; Schlenker & Lobell, 2010), less is known about how much these estimates depend on the EO product used. We hypothesised that adverse weather reduces agricultural productivity (H1a) and rural welfare (H1b). We operationalised weather shocks as departures from each country's normal climate in two independent EO products. To the best of our knowledge, this is the first study to measure EO disagreement for all SSA countries over 25 years and to link it directly to the estimated effects of weather on agriculture.

Our results show a positive relationship between rainfall and agricultural productivity and a negative relationship between heat and cereal yields (H1a), but the story is more nuanced for welfare (H1b). The productivity results are in line with evidence that rainfall drives growth in African agriculture (Barrios et al., 2010) and that high temperatures reduce yields of staple crops (Lobell et al., 2011; Schlenker & Roberts, 2009). The size of our corrected estimate, a 22.6 percent yield loss for a 1 °C warmer year, is large but consistent with crop models and field studies showing steep losses when temperatures exceed crop tolerances (Schlenker & Lobell, 2010).

A closer examination of the welfare results suggests that the link between weather and welfare is hard to see at the national level. Official crop production is associated with lower undernourishment, which supports the view that harvests matter for food security (Hill & Porter, 2017). But when we use only the weather-driven part of crop production, the effect disappears. This may be because national welfare indicators, such as the FAO undernourishment estimate and the UN under-five mortality estimate, are modelled from surveys that occur every few years, so they change smoothly and cannot capture year-to-year shocks (Deaton, 2005). It may also be because households and governments smooth consumption through savings, transfers and imports (Dercon, 2004).

This pattern helps explain why many studies find clear weather effects on output but weaker effects on welfare. The problem of measurement applies to outcomes as well as to shocks: our finding that official cereal yields have been revised by up to 12 percent on average for some periods shows that official agricultural statistics also contain large errors (Carletto, Jolliffe, & Banerjee, 2015; Jerven, 2013). When both sides of the equation are measured with error, true effects become hard to detect.

With those results, our study provides insights into how weather shapes agriculture in SSA. These results are important because they suggest that the damage caused by weather shocks has been understated in studies that rely on a single EO product, and that the true sensitivity of African agriculture to rainfall and heat is about two to six times larger than single-product estimates imply.

5.2. Measurement error and statistical capacity

EO measurement error is seen as a small technical issue in much of the applied literature. Our results show that it is large and consequential (H2a). The two products we study disagree on the sign of the rainfall anomaly in more than four of every ten country-years, and correcting for this disagreement increases the estimated effects of rainfall on crop output, food output and farm labour productivity by between two and six times. These results support the econometric view that noisy measures of explanatory variables lead to large attenuation (Bound et al., 2001; Griliches & Hausman, 1986) and the finding of Michler et al. (2022) that EO product choice changes weather estimates in household data.

Statistical capacity moderates the size of measurement error (H2b). Disagreement between products is larger in countries with weak statistical capacity, low statistical performance and political instability, and single-product estimates are more strongly attenuated there. This may be because both gridded station products and reanalyses rely on national weather services for ground observations, and these services are weakest where statistical systems are weakest (Harris et al., 2020; Hersbach et al., 2020). Once measurement error is corrected, the apparent difference in weather sensitivity between high- and low-capacity countries largely disappears, which suggests that farming in low-capacity countries is not less sensitive to weather; it is simply measured worse.

The growing disagreement over time is also revealing. ERA5 shows rainfall below normal throughout the period and faster warming, while CRU TS shows rainfall at or above normal and slower warming. The fall in the within-country correlation of rainfall anomalies from 0.63 to 0.41 may reflect the continued decline of reporting stations in parts of Africa, which leaves CRU TS more dependent on interpolation and long-term averages, together with changes in the satellite inputs used by ERA5 (Harris et al., 2020; Hersbach et al., 2020). Whatever the cause, it means that recent years, which matter most for current policy, are measured less consistently than earlier ones.

However, to fully interpret these results, we must situate them within SSA's institutional context. Weather services and statistical offices in many countries are underfunded, and their data are often not shared internationally (Devarajan, 2013; Jerven, 2013). EO products were meant to fill this gap, but our results show that they partly inherit it. In countries where ground observations are scarce, satellites and models cannot be checked against reality, so their errors remain hidden unless two products are compared.

6. Theoretical contributions

This study makes three contributions to the development economics literature. First, we advance the literature on EO data by highlighting the dual nature of satellite and reanalysis products. Previous research has mainly focused on the benefits of EO data for measuring weather, yields and poverty (e.g., Burke et al., 2021; Jean et al., 2016). Our longitudinal data and findings, however, suggest that while EO data are a crucial source of information, they also contain large and growing measurement error, which can hide or even reverse estimated effects.

Indeed, we begin to uncover the potential dark side of relying on a single EO product. While EO data can offer consistent coverage, they can also expose researchers to risks such as attenuation bias (Bound et al., 2001), sign reversals across products and vintages (Auffhammer et al., 2013), and false conclusions about differences between countries. This is an important theoretical implication because it reveals a trade-off that demands theoretical refinement: scholars must treat EO data as measurements with error, whose quality depends on the ground data that anchor them, rather than as direct observations of reality.

Second, similar to past studies that have shown that measurement error distorts estimates in African agriculture (e.g., Abay et al., 2019; Gollin & Udry, 2021), our study goes a step further by measuring the error in EO weather data directly and correcting it. We show that the attenuation is large, and that correcting it with a second independent product is a practical solution that any researcher can apply. This contributes to the measurement error literature by applying the errors-in-variables approach of Griliches and Hausman (1986) within a second-generation panel framework that allows for cross-sectional dependence and heterogeneous slopes (Pesaran, 2006).

Third, we contribute to the literature on statistical capacity by showing that EO measurement error is systematically related to the strength of national data systems. The few existing studies of statistical

capacity focus on official statistics (e.g., Dang et al., 2023; Devarajan, 2013). By linking the disagreement between EO products to statistical capacity and political stability, we show that the data gap between countries extends to data collected from space.

Uncovering empirically how EO measurement error affects research findings is also important for policy. First, researchers and agencies that use EO data should compare at least two independent products, report how results change, and use one product as an instrument for the other when estimating weather effects. Second, governments and development partners should invest in national weather stations and statistical systems, because EO products are only as good as the ground observations that anchor them; this includes restoring closed stations, sharing station data with international data centres and supporting African meteorological services. Third, estimates of climate damage used to plan adaptation, insurance and social protection should be corrected for measurement error, since single-product estimates understate the sensitivity of African agriculture to rainfall and heat. Fourth, index-based insurance and early warning systems should be designed with the knowledge that products disagree, for example by combining products or calibrating them with local data, to reduce basis risk for farmers. Fifth, statistical agencies should publish revision histories for agricultural statistics and invest in high-frequency welfare measurement, so that the effects of shocks on welfare can be observed rather than modelled. Finally, support should target low-capacity and unstable countries first, where measurement error is largest and where the gap between measured and true shocks is most likely to mislead policy.

7. Limitations and future research

While this study used a comprehensive dataset from recognised sources, its focus was on country-level averages, which hide large differences within countries and cannot show how measurement error affects individual farmers. We compared two EO products; other products, such as CHIRPS and satellite soil moisture, may have different errors. Our correction assumes that the errors of ERA5 and CRU TS are unrelated; if they share station inputs, our corrected estimates are lower bounds of the true effects. The welfare indicators we used are modelled and smooth, which limits our ability to detect year-to-year welfare effects, and our statistical capacity measure is available only for part of the period.

Future research could expand on these preliminary insights. For example, future work could compare more products at the level of districts or households, using the LSMS-ISA surveys and other geo-referenced data, to measure how EO errors vary within countries. Further research can examine the causes of the growing disagreement between products, such as the decline in reporting stations and changes in satellite inputs. It will also be important to study how measurement error in EO data affects the design and performance of index insurance, drought early warning and targeting of social protection. Finally, future studies could combine EO data with high-frequency phone surveys to measure welfare responses to weather shocks more accurately.

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. *Journal of Development Economics, 139*, 171–184. https://doi.org/10.1016/j.jdeveco.2019.03.008

Auffhammer, M., Hsiang, S. M., Schlenker, W., & Sobel, A. (2013). Using weather data and climate model output in economic analyses of climate change. *Review of Environmental Economics and Policy, 7*(2), 181–198. https://doi.org/10.1093/reep/ret016

Barrios, S., Bertinelli, L., & Strobl, E. (2010). Trends in rainfall and economic growth in Africa: A neglected cause of the African growth tragedy. *Review of Economics and Statistics, 92*(2), 350–366. https://doi.org/10.1162/rest.2010.11212

Bound, J., Brown, C., & Mathiowetz, N. (2001). Measurement error in survey data. In J. J. Heckman & E. Leamer (Eds.), *Handbook of econometrics* (Vol. 5, pp. 3705–3843). Elsevier. https://doi.org/10.1016/S1573-4412(01)05012-7

Burke, M., Driscoll, A., Lobell, D. B., & Ermon, S. (2021). Using satellite imagery to understand and promote sustainable development. *Science, 371*(6535), eabe8628. https://doi.org/10.1126/science.abe8628

Carletto, C., Gourlay, S., & Winters, P. (2015). From guesstimates to GPStimates: Land area measurement and implications for agricultural analysis. *Journal of African Economies, 24*(5), 593–628. https://doi.org/10.1093/jae/ejv011

Carletto, C., Jolliffe, D., & Banerjee, R. (2015). From tragedy to renaissance: Improving agricultural data for better policies. *Journal of Development Studies, 51*(2), 133–148. https://doi.org/10.1080/00220388.2014.968140

Chudik, A., Mohaddes, K., Pesaran, M. H., & Raissi, M. (2016). Long-run effects in large heterogeneous panel data models with cross-sectionally correlated errors. In G. González-Rivera, R. C. Hill, & T.-H. Lee (Eds.), *Essays in honor of Aman Ullah* (Advances in Econometrics, Vol. 36, pp. 85–135). Emerald. https://doi.org/10.1108/S0731-905320160000036013

Chudik, A., & Pesaran, M. H. (2015). Common correlated effects estimation of heterogeneous dynamic panel data models with weakly exogenous regressors. *Journal of Econometrics, 188*(2), 393–420. https://doi.org/10.1016/j.jeconom.2015.03.007

Dang, H.-A. H., Pullinger, J., Serajuddin, U., & Stacy, B. (2023). Statistical performance indicators and index—a new tool to measure country statistical capacity. *Scientific Data, 10*, 146. https://doi.org/10.1038/s41597-023-01971-0

Deaton, A. (2005). Measuring poverty in a growing world (or measuring growth in a poor world). *Review of Economics and Statistics, 87*(1), 1–19. https://doi.org/10.1162/0034653053327612

Dell, M., Jones, B. F., & Olken, B. A. (2012). Temperature shocks and economic growth: Evidence from the last half century. *American Economic Journal: Macroeconomics, 4*(3), 66–95. https://doi.org/10.1257/mac.4.3.66

Dell, M., Jones, B. F., & Olken, B. A. (2014). What do we learn from the weather? The new climate-economy literature. *Journal of Economic Literature, 52*(3), 740–798. https://doi.org/10.1257/jel.52.3.740

Dercon, S. (2004). Growth and shocks: Evidence from rural Ethiopia. *Journal of Development Economics, 74*(2), 309–329. https://doi.org/10.1016/j.jdeveco.2004.01.001

Desiere, S., & Jolliffe, D. (2018). Land productivity and plot size: Is measurement error driving the inverse relationship? *Journal of Development Economics, 130*, 84–98. https://doi.org/10.1016/j.jdeveco.2017.10.002

Devarajan, S. (2013). Africa's statistical tragedy. *Review of Income and Wealth, 59*(S1), S9–S15. https://doi.org/10.1111/roiw.12013

Driscoll, J. C., & Kraay, A. C. (1998). Consistent covariance matrix estimation with spatially dependent panel data. *Review of Economics and Statistics, 80*(4), 549–560. https://doi.org/10.1162/003465398557825

Eberhardt, M., & Teal, F. (2011). Econometrics for grumblers: A new look at the literature on cross-country growth empirics. *Journal of Economic Surveys, 25*(1), 109–155. https://doi.org/10.1111/j.1467-6419.2010.00624.x

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. *Scientific Data, 2*, 150066. https://doi.org/10.1038/sdata.2015.66

Gollin, D., & Udry, C. (2021). Heterogeneity, measurement error, and misallocation: Evidence from African agriculture. *Journal of Political Economy, 129*(1), 1–80. https://doi.org/10.1086/711369

Griliches, Z., & Hausman, J. A. (1986). Errors in variables in panel data. *Journal of Econometrics, 31*(1), 93–118. https://doi.org/10.1016/0304-4076(86)90058-8

Harris, I., Osborn, T. J., Jones, P., & Lister, D. (2020). Version 4 of the CRU TS monthly high-resolution gridded multivariate climate dataset. *Scientific Data, 7*, 109. https://doi.org/10.1038/s41597-020-0453-3

Hausman, J. (2001). Mismeasured variables in econometric analysis: Problems from the right and problems from the left. *Journal of Economic Perspectives, 15*(4), 57–67. https://doi.org/10.1257/jep.15.4.57

Henderson, J. V., Storeygard, A., & Weil, D. N. (2012). Measuring economic growth from outer space. *American Economic Review, 102*(2), 994–1028. https://doi.org/10.1257/aer.102.2.994

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., et al. (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society, 146*(730), 1999–2049. https://doi.org/10.1002/qj.3803

Hill, R. V., & Porter, C. (2017). Vulnerability to drought and food price shocks: Evidence from Ethiopia. *World Development, 96*, 65–77. https://doi.org/10.1016/j.worlddev.2017.02.025

Hsiang, S. (2016). Climate econometrics. *Annual Review of Resource Economics, 8*, 43–75. https://doi.org/10.1146/annurev-resource-100815-095343

Jean, N., Burke, M., Xie, M., Davis, W. M., Lobell, D. B., & Ermon, S. (2016). Combining satellite imagery and machine learning to predict poverty. *Science, 353*(6301), 790–794. https://doi.org/10.1126/science.aaf7894

Jerven, M. (2013). *Poor numbers: How we are misled by African development statistics and what to do about it*. Cornell University Press.

Kapetanios, G., Pesaran, M. H., & Yamagata, T. (2011). Panels with non-stationary multifactor error structures. *Journal of Econometrics, 160*(2), 326–348. https://doi.org/10.1016/j.jeconom.2010.10.001

Kaufmann, D., Kraay, A., & Mastruzzi, M. (2011). The Worldwide Governance Indicators: Methodology and analytical issues. *Hague Journal on the Rule of Law, 3*(2), 220–246. https://doi.org/10.1017/S1876404511200046

Lobell, D. B., Azzari, G., Burke, M., Gourlay, S., Jin, Z., Kilic, T., & Murray, S. (2020). Eyes in the sky, boots on the ground: Assessing satellite- and ground-based approaches to crop yield measurement and analysis. *American Journal of Agricultural Economics, 102*(1), 202–219. https://doi.org/10.1093/ajae/aaz051

Lobell, D. B., Schlenker, W., & Costa-Roberts, J. (2011). Climate trends and global crop production since 1980. *Science, 333*(6042), 616–620. https://doi.org/10.1126/science.1204531

Michler, J. D., Josephson, A., Kilic, T., & Murray, S. (2022). Privacy protection, measurement error, and the integration of remote sensing and socioeconomic survey data. *Journal of Development Economics, 158*, 102927. https://doi.org/10.1016/j.jdeveco.2022.102927

Pesaran, M. H. (2006). Estimation and inference in large heterogeneous panels with a multifactor error structure. *Econometrica, 74*(4), 967–1012. https://doi.org/10.1111/j.1468-0262.2006.00692.x

Pesaran, M. H. (2007). A simple panel unit root test in the presence of cross-section dependence. *Journal of Applied Econometrics, 22*(2), 265–312. https://doi.org/10.1002/jae.951

Pesaran, M. H. (2015). Testing weak cross-sectional dependence in large panels. *Econometric Reviews, 34*(6–10), 1089–1117. https://doi.org/10.1080/07474938.2014.956623

Pesaran, M. H., & Yamagata, T. (2008). Testing slope homogeneity in large panels. *Journal of Econometrics, 142*(1), 50–93. https://doi.org/10.1016/j.jeconom.2007.05.010

Schennach, S. M. (2016). Recent advances in the measurement error literature. *Annual Review of Economics, 8*, 341–377. https://doi.org/10.1146/annurev-economics-080315-015058

Schlenker, W., & Lobell, D. B. (2010). Robust negative impacts of climate change on African agriculture. *Environmental Research Letters, 5*(1), 014010. https://doi.org/10.1088/1748-9326/5/1/014010

Schlenker, W., & Roberts, M. J. (2009). Nonlinear temperature effects indicate severe damages to U.S. crop yields under climate change. *Proceedings of the National Academy of Sciences, 106*(37), 15594–15598. https://doi.org/10.1073/pnas.0906865106

Sen, A. (1981). *Poverty and famines: An essay on entitlement and deprivation*. Oxford University Press.

Stock, J. H., & Yogo, M. (2005). Testing for weak instruments in linear IV regression. In D. W. K. Andrews & J. H. Stock (Eds.), *Identification and inference for econometric models: Essays in honor of Thomas Rothenberg* (pp. 80–108). Cambridge University Press.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- measuringwhatmattersAppendix.docx