Variable Selection in Economic Applications of Remotely Sensed Weather Data: Evidence from the LSMS-ISA

VARIABLE SELECTION IN ECONOMIC APPLICATIONS OF REMOTELY SENSED WEATHER DATA: EVIDENCE FROM THE LSMS-ISA

by

Chandrakant Dipak Agme

A Thesis Submitted to the Faculty of the

DEPARTMENT OF AGRICULTURAL AND RESOURCE ECONOMICS

In Partial Fulfillment of the Requirements

For the Degree of

MASTER OF SCIENCE

In the Graduate College

THE UNIVERSITY OF ARIZONA

2024

THE UNIVERSITY OF ARIZONA
GRADUATE COLLEGE

As members of the Master's Committee, we certify that we have read the thesis prepared by: Chandrakant Dipak Agme
titled:

and recommend that it be accepted as fulfilling the thesis requirement for the Master's Degree.

Date: Jul 28, 2024
Anna Josephson

Date: Jul 31, 2024
Jeffrey Michler

George Frisvold
George Frisvold (Jul 31, 2024 11:18 PDT)
Date: Jul 31, 2024
George Frisvold

Final approval and acceptance of this thesis is contingent upon the candidate's submission of the final copies of the thesis to the Graduate College.

We hereby certify that we have read this thesis prepared under our direction and recommend that it be accepted as fulfilling the Master's requirement.

Date: Jul 28, 2024
Anna Josephson
Thesis Committee Co-Chair
Department of Agricultural and Resource Economics

Date: Jul 31, 2024
Jeffrey Michler
Thesis Committee Co-Chair
Department of Agricultural and Resource Economics

ACKNOWLEDGMENTS

I want to express my sincere gratitude to my advisors, Dr. Anna Josephson, Dr. Jeffrey Michler, and Dr. George Frisvold, for their unwavering support, understanding, and guidance. I am truly thankful.

I would also like to extend my thanks to the Department of Agricultural and Resource Economics and all the faculty and staff for their assistance and support.

Finally, I want to show my appreciation to my friends and family, without whom this journey would not have been possible.

LAND ACKNOWLEDGMENT

We respectfully acknowledge the University of Arizona is on the land and territories of Indigenous peoples. Today, Arizona is home to 22 federally recognized tribes, with Tucson being home to the O'odham and the Yaqui. Committed to diversity and inclusion, the University strives to build sustainable relationships with sovereign Native Nations and Indigenous communities through education offerings, partnerships, and community service.

DEDICATION

To the Department of Agricultural and Resource Economics.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

The rise in the availability of remotely sensed weather data has resulted in economists predicting different outcomes using rainfall as an explanatory or instrumental variable (IV). We analyze 174 papers to identify common rainfall metrics used as an instrument and show the extent of their ad hoc use in predicting a range of outcomes. We use agricultural productivity as a case study to examine the suitability of using different rainfall metrics as an IV. To that extent, we test the predictive power of the 14 most common rainfall metrics in the economics literature, calculated through six remote sensing products across six countries, on agricultural productivity. We find a large amount of heterogeneity in the performance of rainfall metrics. We also find concerning evidence about the validity of using rainfall metrics as an instrument, especially regarding possible exclusion restriction violations and weak instrument problems. Our findings emphasize the need for researchers to carefully select and justify their use of a particular rainfall metric to improve the reliability of their analysis.

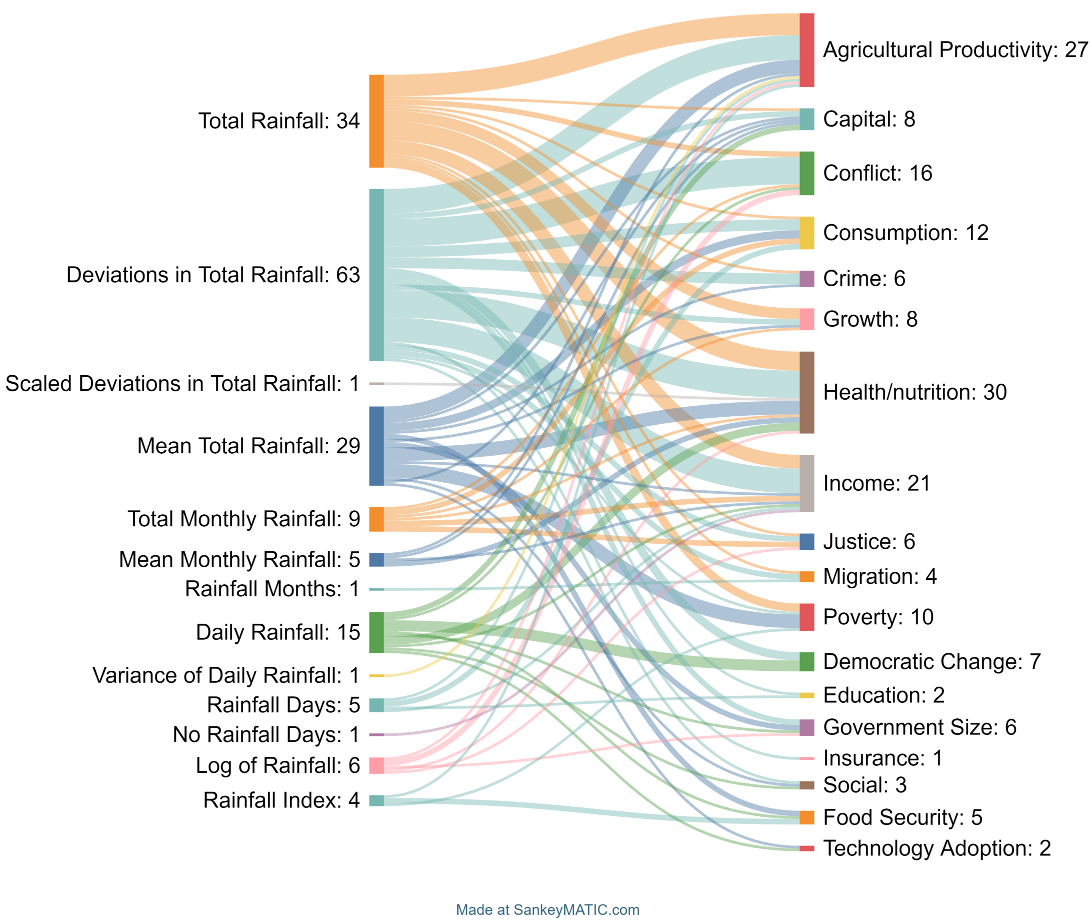
CHAPTER 1

Introduction

The wide availability of remote sensing weather data has led to an increase in the number of economists using different types of rainfall metrics to predict multiple outcomes. However, this use occurs in an ad hoc manner. These economists use rainfall as an explanatory variable or an instrument to predict outcomes from capital to conflict to consumption. The haphazard use of rainfall comes with the backdrop of little guidance on how, when, and where specific rainfall metrics could be used. To the best of our knowledge, this paper is the first to provide guidance on variable selection in the use of remote sensing data integrated with socio-economic data.

We provide concrete evidence on the extent of the ad hoc nature of the use of multiple rainfall metrics to look at different outcome categories in the available economics literature. We review 174 papers that use weather as an instrumental variable. These findings are presented in Figure 1.1. This "Sankey" diagram clearly shows the unsystematic nature of researchers' use of rainfall metrics. We see that papers use multiple rainfall metrics like total rainfall, deviations in total rainfall, and mean annual rainfall to predict agricultural productivity. Should researchers use one over another, or are all of the rainfall metrics substitutes and do not impact the results of a researcher's findings? Looking at specific examples of papers in our sample, we see that to predict agricultural productivity, Banerjee and Iyer (2005) use mean annual rainfall, whereas Hughes (2011) use mean monthly rainfall. Similarly, Jacoby and Skoufias (1998) use rainfall days, but Ghimire Monika et al. (2016) use total monthly rainfall to understand consumption. Also, to predict health and nutrition, Hanandita and Tampubolon (2014) use the deviations in total rainfall, whereas Mapulanga and Naito (2019) use the log of rainfall.

Figure 1.1: Rainfall Outcome Relationship



Note: This figure captures relationships between what rainfall variables are used to understand what outcome category. 75 Flows between 31 Nodes. Number of Papers = Total Inputs = Total Outputs = 174.

This paper provides guidance to researchers looking to take a systematic approach to variable selection when using weather data in economic applications. To do this, we calculate fourteen rainfall metrics commonly used in the economics literature. To ensure the broad applicability of our findings, we calculate each metric for six different remote-sensing products favored by economists. Finally, we combine the weather metrics for each remote-sending product with household survey data collected as part of the World Bank Living Standards Measurement Study - Integrated Surveys on Agriculture (LSMS-ISA) in Ethiopia, Malawi, Niger, Nigeria, Tanzania and Uganda. We test the predictive power of each individual rainfall metrics-remote sensing pair on agricultural productivity, a common first-stage IV regression. Intuitively, if all these different measures of rainfall are equally valid instruments, then each should be a significant predictor of agricultural productivity (though signs and magnitudes will vary depending on the specific metric). We look for metrics that (a) are consistently significant across remote sensing products and countries, (b) have a consistent sign, and (c) are similar in significance and sign to other metrics, which would suggest a degree of substitutability between the metrics.

In addition to the concerns about the ad hoc choice of weather metrics for IVs, there is a long-simmering concern about whether weather even satisfies the criteria for a valid instrument. In short, does weather pass the exclusion restriction? Deaton (2010) draws a distinction between a variable that is *external*, whose value is determined outside the system, and a variable that is *exogenous*, orthogonal to the error term. Weather is clearly the former but, depending on the context of the research question, may not be the latter. Taking an empirical approach, Sarsons (2015) shows that in the context of conflict stemming from income shocks in India, rainfall shocks, a frequently used IV, fail to pass the exclusion restriction. We examine if rainfall in our setting is exogenous using a straightforward heuristic approach. Intuitively, if rainfall is exogenous, it should be uncorrelated with household unobservables. Therefore, the significance of any given rainfall metric in predicting agricultural productivity should remain relatively constant in regressions with and without household fixed effects. If this is not the case, it suggests that rainfall is correlated with household unobservables that, absent panel data, end up in the error term.

There are three headline findings from our research. First, there is a large amount of heterogeneity in the way rainfall metrics perform across remote sensing products and across countries. Few rainfall metrics are consistently significant and have a consistent sign. Viewing our regressions of agricultural productivity on rainfall as a first stage IV regression, our results show that, depending on the source of the rainfall data and the country of study, the same rainfall metric might be a weak instrument (not significant) or might be an inconsistent predictor of agricultural productivity (significant but with opposite signs).

Our second headline finding is that the inconsistency in rainfall as a predictor of agricultural productivity means that few metrics are good substitutes for each other. As Figure 1.1 shows, economists operate with the implicit assumption that many weather metrics are substitutes for each other when looking to instrument for the same outcome variable. Exceptions to this lack of substitutability are variables that are calculated as deviations from some measure of rainfall. Examples include total seasonal rainfall and deviations in total seasonal rainfall as well as the percentage of rainy days in a season and its deviation.

Our final headline finding is that in regressions that lack household fixed effects (e.g., time-invariant household unobservables are captured in the error term), rainfall metrics are significant predictors of agricultural productivity 80% of the time. Once we include household fixed effects, removing household unobservables from the error term, rainfall metrics are significant predictors of agricultural productivity only 35% of the time. We interpret this as evidence that, while rainfall is external to some household decision-making, it is correlated with time-invariant, unobservable characteristics of the household, like their past experience with weather shocks or their geographic location. Absent an ability to control for household fixed effects, rainfall fails the exclusion restriction. However, having controlled for household fixed effects, rainfall has a weak instrument problem.

We contribute to the growing body of empirical research examining instrumental variables, their validity, exogeneity, and potential weakness. Violations of the exclusion restriction and the problem of weak instruments, as well as solutions to these problems, is well documents in the theoretical literature (Angrist and Imbens, 1995; Stock and Yogo, 2002; Andrews et al., 2019; Kiviet, 2020; Mellon, 2023). Following on the critique in Deaton (2010), applied economists turned their attention to how certain instruments perform in specific empirical contexts. Sarsons (2015) finds that rainfall shocks significantly impact income in areas upstream of dams in India but not downstream of the dams. Despite this, rainfall shocks remain strong predictors of riots. The conclusion is that the strength of rainfall as an instrument varies by geographic location and that it impacts riots through a channel independent of income. Mellon (2023) explores the many alternative channels through which rainfall might impact the outcome variable, violating the exclusion restriction. He analyzes 195 papers and finds that "the underlying assumptions of weather-IVs are not strictly true, and many results are likely wrong." While our research question is the same as Sarsons (2015) and Mellon (2023), namely when and what rainfall variables are valid instruments, we take a different approach. We focus exclusively on how rainfall performs in a first-stage regression and make recommendations on which rainfall metrics are strong instruments, have a consistent sign, and are substitutable with other metrics.

The implications of our findings speak to the extremely large literature that relies upon weather as an instrument to identify causal effects. Just since 2020, this includes papers using mean annual rainfall (Nasser et al., 2022), deviations in total rainfall (Georgiadis et al., 2021; Trinh et al., 2021, 2022), and total rainfall (Faradiba, 2021) all to predict health and nutrition. Researchers use deviations in total rainfall to predict agricultural productivity (Veljanoska, 2022), migration (Palacios and Pérez-Uribe, 2021), welfare (Ngoma et al.), and capital (Kalemli-Özcan et al., 2020). Others use mean annual rainfall to predict capital (Kling et al., 2021), poverty (Heger et al., 2020), and consumption (Hassan, 2020). As we have said before, this ad hoc approach is dominant but not universal. As an example, Munley et al. (2023), Lind (2020), and Rudolph (2020) all use daily rainfall to predict democratic change. But even if daily rainfall is frequently use as an IV in the context of studying democratic change, its use to predict other outcomes by other researchers suggests that it fails the exclusion restriction for someone (Mellon, 2023). Our findings speak to this literature as we provide insights regarding whether researchers should be using one weather metrics over others and how their results might change if they had chosen a different metric or a different source of weather data. Also, climate and weather variables are complex instrumental variables because they are either invariant or nearly invariant over time. The problems we uncover would be common to any time-invariant variable. With panel data, this might also be a problem for variables that vary over time but not much over space, which is a spatial but not a temporal problem.

The remaining paper is organized as follows. In Chapter 2 we discuss how the existing economics literature uses weather as an instrumental variables and document further the ad hoc nature of choosing a specific weather metric to instrument a wide variety of potentially endogenous variables. In Chapter 3 we provide details on the weather and household data used to generate our first-stage estimates of rainfall on agricultural productivity. We also discuss our pre-specified analysis plan, the econometric models, and our heuristic method of inference. In Chapter 4 we discuss our results in three stages. First, we present descriptive statistics that show the pattern of different rainfall metrics draw from different remote sensing sources. Second, we analyze the share of the significance point estimates for each rainfall variable with and without fixed effects. Third, we look at the coefficients on each rainfall metrics for each country. Finally, we conclude with Chapter 5, giving our recommendations for best practices when a researcher wants to use rainfall as an instrument.

CHAPTER 2

How Economists Use Weather Data

The economics literature uses causal inference techniques to understand the effect of explanatory variables on outcomes. If explanatory variables are endogenous, researchers use the instrumental variables (IV) technique. Many studies incorporating IV use rainfall or other weather metrics as an instrument (Dell et al., 2014). Thus, to understand how the current economics literature looks at and incorporates various rainfall metrics as an instrument, we generate a dataset that captures the universe of economic papers that use rainfall as an IV. We then use ChatGPT to "read" and categorize these papers based on the endogenous rainfall variables as an instrument. Finally, we discuss the patterns of their use in this body of literature.

2.1 Generating the Literature Data Set

To understand economists' use of rainfall as an IV, we first look at the current economics literature that employs rainfall as an instrument. We examine the various scholarly sources from which to scope academic literature. This scoping exercise included Google Scholar, Web of Science, JSTOR, PubMed, Science.gov, Scopus, ResearchGate, and OpenAlex. We select OpenAlex (Priem et al., 2022), named after the Library of Alexandria, for three reasons. First, OpenALex is the successor to the Microsoft Academic Graph (MAG), which was retired on December 31, 2021. The MAG was "a heterogeneous graph containing scientific publication records, citation relationships between those publications, as well as authors, institutions, journals, conferences, and fields of study" (Sinha et al., 2015). OpenAlex's database includes sources like the MAG, Crossref, ORCID, ROR, DOAJ, Unpaywall, Pubmed, Pubmed Central, the ISSN International Center, and other institutional repositories like arXiv. The MAG has been used for meta-analysis of research trends in economics by (Jones, 2021; Josephson and Michler, 2023). It indexes 248 million academic works (nearly three times of Scopus and Web of Science), making it a very diverse platform of research works (Priem et al., 2022). Second, OpenAlex is not just a search tool or a database, but like the MAG, it is a catalog disambiguating connecting scholarly work. For each work, metadata such as the title of the research paper, the author, and the year the paper was published are created, as well as concepts that capture what the work is about. OpenAlex indexes sixty-five thousand concepts. Third, OpenAlex is an open, free-to-use catalog that allows queries via API. This not only directly allows us to contribute to open science but indirectly as it allows us to increase the reproducibility of our work.

To capture papers relevant to this study, we use Boolean terms for relevant concepts. Our query included "((Weather) AND (Instrumental Variable)) OR ((Rainfall) AND (Instrumental Variable))". This resulted in more than 65,000 results. After filtering for only English language papers and papers in "Economics", "Econometrics", and "Finance", the search results narrowed to 3,062 papers.[1] (Priem et al., 2022) These papers comprise what is, in our view, the potential universe of economic papers that use rainfall as an IV. For each paper we have, OpenAlex provides complete bibliographic information and, based on our search criteria, citation count, an OpenAlex location ID, authors, affiliations, and the number of times the paper has been cited. With the output from OpenAlex, we began with the first 300 research works (10% of the total) according to the relevance score produced by OpenAlex.

2.2 "Reading" the Literature

With the metadata on papers generated from OpenAlex, we aim to parse through these papers efficiently so that we can better understand how the economic literature uses rainfall as an instrument. Reading all 3,062 papers would be a lengthy task. To speed up the process

[1] One shortcoming of OpenAlex is that other researchers working with OpenAlex may get a different number of papers, even using the same criteria and search process. OpenAlex is continuously adding more papers, and/or the papers are cited more, both of which change the "relevance score" of papers. According to OpenAlex: "The relevance_score is based on text similarity to your search term. It also includes a weighting term for citation counts: more highly-cited entities score higher, all else being equal."

and to create a road map for other researchers attempting to parse a large body of literature, we use ChatGPT to do the "reading".

ChatGPT (Chat-Generative Pre-Trained Transformer 4.0) allows us to "read" each paper and extract outcome variables, explanatory variables, the presence of weather instruments, and rainfall variables into a .csv file. Our decision to use ChatGPT 4.0 by OpenAI over other artificial intelligence models like Bard by Google, Perplexity AI, or others, stems from the current understanding of ChatGPT 4.0 to be the most efficient, widely available, and largest natural language processing (NLP) model (as of this writing).

To optimize our GPT model for reading research papers successfully and obtaining our desired information, we created a custom GPT called Econ Analyst. We told Econ Analyst GPT that its purpose was to "analyzes hundreds of economics papers in PDF format for IV usage and structure." We thus generate prompts to help get consistent information about the papers and tabulate the results in order to shorten this workflow from OpenAlex to analyzable data. Econ Analyst GPT's instructions are to:

> Analyze hundreds of economics research papers in PDF format, specifically focusing on the use of Instrumental Variables (IV). When reviewing a paper, [it should] identify whether an IV is present or not and provide a brief analysis of the paper's structure, including the dependent variable (y), the explanatory variables (x), the instrumental variable (if present), and the variable that the instrument is replacing (z). Additionally, [it should] mention the source of the data used in the research. In [its] responses, [it should] clearly indicate with a '1' if an IV is present or a '0' if not. [Its] goal is to assist users in understanding the application and presence of IV in economic research.

We then upload a batch of ten papers and provide the following additional instructions:

> Please analyze the entire PDF research paper with its title, focusing on the presence of Instrumental Variables (IV), and provide the information in a list format: 1. IV (yes_no) – indicate with a '1' if IV is present, '0' if not. 2. y – identify the dependent variable. 3. x – list the explanatory variables. 4. z – specify the instrument being used. 5. source – mention the data source. Also,

> detail the type of rainfall or temperature variable used, such as annual rainfall or mean daily temperature.

We arrived at these instructions after numerous iterations. Multiple versions of the instruction were trialed, and when Econ Analyst GPT gave a reasonable answer that was as close to what we were looking for, we asked Econ Analyst GPt, "How should we prompt you so that you'll produce the results you just produced." Iterations in this fashion eventually led to the instructions given above.

After ten papers were analyzed by Econ Analyst GPT, we asked it to convert the generated data into a more accessible tabular format. We chose to analyze ten papers before asking them to tabulate, as sometimes, with more than ten papers, Econ Analyst GPT would claim it couldn't read PDFs or would analyze one or two papers, duplicate the results till it had its output, then claim it had read all the ten individual papers. As with the instructions to read the paper, when we got satisfactory output data, we asked Econ Analyst GPT what we should ask it so that it would produce the results output data. This instruction given by Econ Analyst GPT was:

> Please convert the analyzed PDFs into a tabular format in the order I submitted them with the following columns: 'Paper,' 'IV (yes_no),' 'Dependent Variable,' 'Explanatory Variables,' 'Instrument,' 'Data Source,' and 'Type of Rainfall/Temperature Variable.' Each PDF should be represented in one row of the table.

The use of ChatGPT allows us to "read" many more papers in a shorter time than a human could. That said, our GPT required a great deal of revision and quality checks. Once we had data from 10% sample of 3,062 papers filed into a .csv file, we performed a verification check to ensure that ChatGPT gave us accurate information. We opened a selection of papers analyzed by Econ Analyst GPT and confirmed if the dependent variable given was correct or not and if it was given a rainfall variable, it was specific to our need or not. During this process, we also categorized outcome variables into one of eighteen categories:

Table 2.1: Categories of Outcome Variables

Outcome	Definition
Agricultural productivity	Using measures of agricultural productivity, e.g. yield, value, technology adoption.
Capital	Using various forms of capital at a macro-level (institutions, state, global).[2]
Conflict	Examining various forms of conflict between nation-states, or different groups.
Consumption	Examining household- or individual-level consumption.
Crime	Using different types of crimes.
Democratic change	Examining change in government institutions.
Education	Examining the impact on education.
Food Security	Examining the determinants of food security.[3]
Government size	Examining the functioning and productivity of various institutions.
Growth	Examining economic growth at a macro level.
Health & nutrition	Examining determinants of health and/or nutrition.
Income	Examining income at an individual and household level.
Insurance	Examining insurance for individuals, households, or businesses.
Justice	Examining the factors around inequality, discrimination, and injustice.
Migration	Examining migration.
Poverty	Examining determinants of poverty.
Social	Examining group behavior, social cohesion, or societal factors.
Technology adoption	Examining at various types of technology adoption.[4]

Note: Categories by the authors, these variables represent the eighteen most common outcomes in the literature.

These variables represent the eighteen most common output categories. We also categorized the rainfall variables into one of the 14 rainfall categories described in Table 3.2. Of the first 300 papers, we found that ChatGPT was unable to or only partially able to analyze 83 papers, as either it could not read the PDF or it gave random output. Of the remaining 217 papers, about 43 were irrelevant to our study as they did not use any rainfall variable and simply mentioned rainfall in the text. Our final sample is 174 papers.

This sample of 174 successfully analyzed papers reveals what rainfall metrics researchers commonly use in their research. The sample also provides us with information about what outcome categories the researchers are trying to instrument for with these rainfall metrics, and sheds more light on the ad hoc matching of rainfall metrics with outcome categories.

2.3 Analysis of the Literature

To understand the unsystematic nature of the use of rainfall metrics by researchers, we look at the Sankey diagram in Figure 1.1, which presents the results from Econ Analyst GPT's "reading" of the papers. In our sample of 174 papers, we see that 36% of the papers use deviations in total rainfall (this is the most used rainfall variable), 20% of the papers use annual rainfall, and 17% of the researchers use mean total rainfall. About 9% of the researchers use daily rainfall. These four rainfall variables account for more than 80% of the rainfall variables used in IVs.

While most papers use one of these four as IV, all instruments for numerous outcomes. The most common of which are health and nutrition (17%), agricultural productivity (15%), income (12%), and conflict (9%). But they are also used to predict everything from consumption (7%), poverty (6%), growth (5%), capital (5%), democratic change (4%), government size (4%), crime and justice (3% each), food security (3%), migration (2%), social (2%), technology adoption (1%), and insurance (1%). Among the 16% papers that look at agricultural productivity, about 30% use annual rainfall, 33% use deviations in total rainfall, and 18% use mean total rainfall. For the 12% of papers looking at income, total rainfall (24%) and deviations in total rainfall (43%) were the most used rainfall variables. However, six other papers use other rainfall metrics to instrument income. Total rainfall (24%) and deviations in total rainfall (33%) are the most used instruments for the 17% of papers that study health and nutrition. But again, six other papers use different measures of rain as an IV. Even for an outcome like democratic change, where researchers exclusively use two weather metrics (deviations in total rainfall or daily rainfall), these two metrics measure distinct types of events like agricultural productivity, capital, and health/nutrition.

When we look at specific papers or researchers, this heterogeneity persists. we see that Brückner (2012) uses annual rainfall to predict growth in one instance, whereas in another instance, Brückner et al. (2020) uses annual rainfall to predict income. Also, Hansford and Gomez (2010) uses daily rainfall to understand democratic change, whereas Fontenla et al. (2019) uses daily rainfall to understand health and nutrition. Similarly, for deviations in total rainfall, Amare et al. (2018) uses it to predict consumption, but Veljanoska (2022) uses deviations in total rainfall to predict agricultural productivity, and Raleigh et al. (2015) uses deviations in total rainfall to predict conflict. As for the log of rainfall, we can see that Brückner and Gradstein (2014) uses that to predict government size, and Owens et al. (2003) uses the log of rainfall to predict agricultural productivity. All the above shows that researchers use the same rainfall variables to predict other things. This should be fine if the same rainfall metric is a good predictor of different outcomes. But is annual rainfall as good a predictor of growth as it is of income, if it is at all?

Based on the review of 174 papers, we see evidence of researchers using rainfall as an instrument in an ad hoc manner. Researchers publishing across economics use different rainfall variables to predict similar outcomes, and/or researchers use the same rainfall variables to predict different outcomes. This is not to say there is no consistency. We see a small amount of consistency, notably in understanding democratic change; researchers use deviations in total rainfall or daily rainfall only. Similarly, in an attempt to understand agricultural productivity, about 80% of the researchers use either total rainfall, deviations in total rainfall, or mean total rainfall. Finally, to predict poverty, we see that researchers predominantly (80%) use total rainfall or mean total rainfall. Given that economists mostly tend to use rainfall as an IV in an unsystematic and ad hoc manner, we aim to understand the implications for these IVs in terms of their strength (relevance) and their ability to pass the exclusion restriction (exogeneity). To that effect, we take agricultural productivity as a test case, as it is one of the most common outcomes, also used as a proxy for well-being, in the economics literature.

CHAPTER 3

Testing the Effects of Various Weather Metrics

Our synthesis of the literature led us to the conclusion that researchers use rainfall metrics in an unsystematic manner. Now, we use agricultural productivity as a test case to understand the effect of using different rainfall variables, generated from multiple remote sensing products on the sign, or significance of estimates. At the same time, we test the validity of two of the four major identification assumptions for an instrumental variable: Stable Unit Treatment Value Assumption (SUTVA); exogeneity of the instrument; relevance of the instrument to the endogenous variable; and monotonicity.

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

3.1 Data

To examine the substitutability of different rainfall metrics and their consistency (the sign and significance), we combine two distinct types of precipitation data and household survey data. The weather data are taken from multiple remote sensing sources, and all are available from 1983. These remote sensing products allow us to do two things. One, calculate multiple rainfall metrics at various levels of aggregation. And two, help us to understand if using one rainfall metric over another, coming from different remote sensing products, will change the sign and significance of the effect on agricultural productivity. To that end, we need household data, which comes from multiple survey instruments from six Sub-Saharan

African countries, part of the World Bank Living Standards Measurement Study - Integrated Surveys on Agriculture (LSMS-ISA).

3.1.1 Weather Data

In selecting weather data, we want to use datasets that are both available in the public domain and extensively used in the applied social sciences. We pre-specified that the data used must: (1) be available at the daily level, which helps us calculate the most commonly available weather metrics; and (2) have at least three decades of data available so we can calculate shocks or deviations from long-term average. This allows us to have uniform temporal resolution and duration to cover all the years from our household data.

Modern remote sensing sources of rainfall data combine information on precipitation by combining satellite data that provides meteorological information at full coverage, rain gauge data that provides site-specific observations, and the outcome of atmospheric reanalysis. Our study includes the following data products: (1) African Rainfall Climatology 2 (ARC2), (2) the Tropical Applications of Meteorology using SATellite data and ground-based observations (TAMSAT), (3) the Climate Hazards group InfraRed Precipitation with Station Data (CHIRPS) (4) the European Centre for Medium-Range Weather Forecasts (ERA5), (5) the NASA Modern-Era Retrospective Analysis for Research and Applications (MERRA-2) and, (6) the NOAA Climate Prediction Center (CPC) Unified Gauge-Based Analysis of Daily Precipitation.

ARC2 uses rain gauge data from the Global Telecommunications System (GTS) and Geostationary Operational Environmental Satellite (GOES) precipitation index (GPI). The data generated from ARC2 are at a resolution of 0.1°. Similarly, TAMSAT uses rain gauge data from the GTS and Meteosat thermal IR. CHIRPS is built on the same approach as ARC2 and TAMSAT but has more climatological products and is at a resolution of 0.05°(Tarnavsky et al., 2014; Novella and Thiaw, 2013; Funk et al., 2015). ERA5 and MERRA-2 are reanalysis data products (Bosilovich et al., 2016; Hennermann and Berrisford, 2020). Assimilation models combine observations gathered from multiple sources to generate a more comprehensive model of the climate system. And reanalysis data products are created are using the assimilation models in such a way that the final product can support climate analysis over a long period of time. CPC created from the information available at the NOAA Climate Prediction Center, and its special collections and GTS is primarily a gauge data product using an Optimal Interpolation (OI) technique at a resolution of 0.5° (Chen et al., 2017). Table 3.1 provides a summary of the characteristics of each remote sensing precipitation product.

To measure precipitation, we use the fourteen rainfall metrics given in Table 3.2. Mean daily rainfall, median, variance of daily rainfall, and skew of the daily rainfall are the moments of daily rainfall distribution for the growing season. Total rainfall and the z-score of total rainfall are cumulative of the daily rainfall. Rainfall days and no rain days are the number of days with at least 1 mm of rain and the number of days with less than 1 mm of rain, respectively. The share of rainy days is the percentage of growing season days with rain. The intr-season dry spell is the maximum length of time in days without rain. Some rainfall metrics also have deviations calculated from the long-run average.

Table 3.1: Sources of Precipitation Weather Data

Dataset	Length of record	Resolution (°)	≈Grid size (km)	Time step	Data	Units
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

Note: The table, adapted from Michler et al. (2019), summarizes the remote sensing sources and related details for precipitation data.

Table 3.2: Weather Variables & Transformations

Rainfall	
Daily rainfall	In mm
Mean	The first moment of the daily rainfall distribution for the growing season
Median	The median daily rainfall for the growing season
Variance	The second moment of the daily rainfall distribution for the growing season
Skew	The third moment of the daily rainfall distribution for the growing season
Total	Cumulative daily rainfall for the growing season
Deviations in total rainfall	Cumulative daily rainfall for the growing season minus the long run average
Scaled deviations in total rainfall	The z-score for cumulative daily rainfall for the growing season
Rainfall days	The number of days with at least 1 mm of rain for the growing season
Deviation in rainfall days	The number of days with rain for the growing season minus the long run average
No rain days	The number of days with less than 1 mm of rain for the growing season
Deviation in no rain days	The number of days without rain for the growing season minus the long run average
Share of rainy days	The percent of growing season days with rain
Deviation in share of rainy days	The percent of growing season days with rain minus the long run average
Intra-season dry spells	The maximum length of time (measured in days) without rain during the growing season

Note: The table presents definitions for included weather variables and transformations from weather sources defined in Michler et al. (2019). Growing season is determined for each country following FAO crop calendar. For variables when "long run" is referenced, long run is defined as the entire length of the weather dataset. While each weather source has a different start date, to ensure a uniform time series all datasets were shortened to 1983, which is the latest start date of the data sources.

3.1.2 Household Data

For the household data, we use the World Bank LSMS-ISA data, taken from surveys designed in collaboration with Sub-Saharan African nations' statistical offices. With support from the Living Standards Measurement Study, this data has multiple rounds of household panel survey data conducted by the statistical offices of Ethiopia, Malawi, Niger, Nigeria, Uganda, and Tanzania for over a decade.

The data for Ethiopia is from the Ethiopia Socioeconomic Survey (ESS) conducted over three rounds of 2011-12, 2013-14, and 2015-16 by the Central Statistical Agency of Ethiopia (Central Statistics Agency of Ethiopia (CSA), 2014, 2015, 2017). The first round of survey data is regionally representative. Whereas the subsequent rounds of surveys added 1,500

Table 3.3: 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	2011	3,968	2,223
	Ménages et l'Agriculture (ECVMA)	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.

households in urban areas. After removing non-agricultural households, the final dataset has 7,272 observations.

The data for Malawi comes the Integrated Household Panel Survey (IHPS), which is longitudinal in nature and is representative at multiple rural-urban and regional-national levels. This data comes from the 2010-11, 2013-14, and 2016-17 rounds of surveys (National Statistical Office (NSO), 2012, 2015, 2017). After removing non-agricultural households and households that relocated, the final dataset has 3,250 observations.

The data from Niger includes two waves: 2011 and 2014 (Survey and Census Division, National Institute of Statistics, Niger (NIS), 2014, 2016). The survey data is representative at multiple rural-urban and regional-national levels. The final dataset, after excluding the non-agricultural households, has 3,913 observations.

The data for Nigeria comes from multiple rounds of General Household Survey data from 2010-11, 2012-13, and 2015-16 (National Bureau of Statistics (NBS), 2012, 2014, 2019). This data, like data from Malawi and Niger, is representative at multiple rural-urban and regional-national levels. The final dataset for three rounds after removing non-agricultural households has 8,384 observations.

We use the Tanzania National Panel Survey (TZNPS) data for Tanzania. This data was collected in three rounds in 2008-09, 2010-11, and 2012-13 (Tanzania National Bureau of Statistics (TNBS), 2011, 2012, 2015). After removing the households that did migrate and non-agricultural households, the final datasets consist of 5,669 observations.

For Uganda, we use the Uganda National Panel Survey (UNPS) data that covers three rounds of data collection from 2009-10, 2010-11, and 2011-12 (Uganda Bureau of Statistics (UBOS), 2019, 2014a,b). Like other countries, the data for Uganda is representative at multiple rural-urban and regional-national levels. The final data set after removing non-agricultural households consists of 5,250 observations.

The data from the six countries are combined to form a single cross-country panel dataset. This combined dataset constitutes 33,738 observations (see Table 3.3). To measure agricultural productivity, we use the yield of the primary crop in kilograms per hectare as well as the total value of the harvest in 2010 US dollars per hectare (see Table 3.4). In some econometric specifications, we include independent variables like labor in number of days per hectare, application of fertilizer in kilograms per hectare, pesticide, herbicide, and irrigation as an indicator variable (equal to 1 if used, 0 otherwise).

Table 3.4: Household Variables and Definitions

***Panel A**: Outcome Variables*	
Yield	Output in kilograms per hectare for the primary cereal crop in each country data set
Value	Output in real USD per hectare for all seasonal farm crop production
***Panel B**: Input Variables*	
Labor use rate	Number of days per hectare
Fertilizer application rate	Kilograms per hectare
Seed application rate	Value in USD per hectare
Pesticide use	Equal to 1 if yes, 0 if no
Herbicide use	Equal to 1 if yes, 0 if no
Irrigation use	Equal to 1 if yes, 0 if no

Note: The table taken from Michler et al. (2019) presents definitions for included outcome and input variables from LSMS sources defined in Table 3.3.

3.2 Analysis Plan

As stated above, the data cleaning, analysis, and the associated results were pre-specified in an OSF pre-analysis plan (Michler et al., 2019).

3.2.1 Estimation Strategy

We follow Deschênes and Greenstone (2007):

$$Y_{ht} = \alpha_h + \gamma_t + \sum_{j}^{J} \beta_j f_j \left(R_{jht}\right) + X_{ht}\pi + u_{ht} \tag{3.1}$$

where Y_{ht} is our outcome variable from the LSMS-ISA-supported household surveys, for household h in year t. X_{ht} is a matrix of input variables from the LSMS-ISA. We include year fixed-effects (γ_t) and control for household fixed-effects (α_h). The function $f_j\left(R_{jht}\right)$ represents our weather variables of interest, where j represents a particular measurement of rainfall. Last, u_{ht} is an idiosyncratic error term clustered at the household level.

From here, we estimate the following three linear models, where for each model, we consider a single rainfall variable:

$$Y_{ht} = \beta R_{ht} + u_{ht} \tag{3.2a}$$

$$Y_{ht} = \alpha_h + \gamma_t + \beta R_{ht} + u_{ht} \tag{3.2b}$$

$$Y_{ht} = \alpha_h + \gamma_t + \beta R_{ht} + X_{ht}\pi + u_{ht} \tag{3.2c}$$

Here, we have many regressions as we include the 14 rainfall variables, six countries, six remote sensing products, and two outcome variables. This gives us a total of 3,024 different regressions [1].

[1] Here, deviating from the pre-analysis plan given by Michler et al. (2019), we do not include eight temperature variables, ten extraction methods, and quadratic specifications. A complete set of results are presented in Michler et al. (2019).

3.2.2 Tests

Unlike a "typical" economics research paper, which would include tables of coefficient estimates, p-values, t-statistics, confidence intervals, and standard errors to show the results, we present our results in a different format. This is motivated as, with more than three thousand regressions, it becomes a challenge to tabulate our results for effective analysis. So, we rely on heuristics that make our analysis more presentable and understandable. The metrics we adopt to showcase our results are (1) the distribution of rainfall metrics originating from different remote sensing products; (2) the share of coefficients of p-values significant at 0.05 with just weather metrics, weather plus fixed effects, and weather plus fixed effects plus input variables and; (3) the coefficient size with 95% confidence intervals for each of the fourteen rainfall metrics by each country.

To compare our metrics across regressions, we apply the following 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.

This approach is built on Levine and Renelt (1992)'s extreme bounds approach and Sala-i Martin (1997)'s graphical methods to visualize the differences.

CHAPTER 4

Results

As mentioned in 3.2.2, to deal with more than three thousand regressions, we present results in a series of figures, which allow us to evaluate the significance, magnitude, and general trends exploring the relationship between rainfall variable selection and its relationship with agricultural productivity. We do this to apply our heuristic criteria so that we can discuss differences across regressions absent test statistics. Our aim is to explain, in a given country, whether rainfall measured via remote sensing products will be a suitable instrument. We begin our discussion by showing that the value of a rainfall metric differs by remote sensing products. Next, we present the share of p-values from individual rainfall variables and observe how that share changes across specifications. We then examine the consistency of the sign and significance of coefficients on rainfall metrics disaggregated by country, remote sensing product, and dependent variable. Finally, we conclude with a discussion of our analysis and draw a set of best practices for researchers.

4.1 Descriptive Statistics

We begin our analysis by examining our fourteen rainfall variables by each remote sensing product in each country. This helps us understand how different rainfall metrics created by various remote sensing products differ from one another and in each country. The between-type variation within a given rainfall variable is due to the varied nature of different remote sensing products. As discussed in Subsection 3.1.1, ARC2, CHIRPS, and TAMSAT use merged gauge and remote sensing data, CPC uses an optimal interpolation technique on gauge data, ERA5 and MERRA-2 are reanalysis data sets. They each have their own data sources, interpretation, and interpolation of precipitation. To make our descriptive analysis more accessible, we take the average value of the metrics for each year in each country and then plot the resulting time series. We do this for all fourteen metrics but limit our discussion in the paper to mean daily rainfall and the share of rainy days.[1]

Figure 4.1 presents the mean of daily rainfall for all countries and all remote sensing products. Considering each country individually, we see that in Ethiopia, CHIRPS does not show much variation, whereas CPC and MERRA-2 show more variation, with large dips from 1999 to 2005. At the same time, CPC, using only gauge data, generally gives mean rainfall values lower than CHIRPS, whereas ERA5, which is reanalysis data, always gives mean rainfall values higher than CHIRPS. These differences can be extreme. In 2011, CPC reported an average of 1 mm of rainfall a day, while ERA5 reports 5 mm of rainfall a day.

Turning to Malawi, there is much more agreement across products than in Ethiopia. However, as in Ethiopia, CPC generally gives mean rainfall values lower than the average, whereas ERA5 always gives mean rainfall values higher than the average. This relative disagreement of products in Ethiopia, compared to Malawi, may be due to some products performing worse in the mountainous terrain of Ethiopia. In Niger, remote sensing products are even more consistent than in Malawi. The only outlier is ERA5, which, while over-reporting rainfall in Ethiopia and Malawi, relative to the other products, under-reports in Niger.

Turning to Nigeria, as in Ethiopia, CHIRPS has relatively low variation, whereas CPC, MERRA-2, ARC2, and TAMSAT show wide variations. MERRA-2 and CPC perform particularly poorly, reporting a daily average rainfall of 11 mm in some years and only 3 mm a day in other years. In some years, the disagreement between products can be pronounced. In 2014, CPC report an average daily rainfall of 3 mm, while ERA5 report 7 mm of rainfall a day.

In Tanzania, like Niger, we see consistent values except for an anomaly in 1999, where CHIPS and ERA5 report over 6 mm of rain a day. After 2000, there is a large agreement between products.

[1]The other 12 rainfall variables and their descriptive figures are available in Appendix A.

In Uganda, like Malawi, we see a fair amount of variation but none of those anomalous spikes that occur in Ethiopia and Nigeria. As in most other countries, CPC tends to report less rainfall than other products, while ERA5 reported more.

Summarizing our findings across countries, there is a broad agreement between CHIRPS, ARC2, TAMSAT, and MERRA-2, while CPC tends to report less rain and ERA5 reports more. But this is not always the case. In Nigeria, CPC and ERA5 did the opposite. And MERRA-2 shows wide variation with implausible spikes and dips.

Next, we turn to Figure 4.2, which presents the share of rainy days by country and by remote sensing product. In Ethiopia, as in mean daily rainfall, we see that CHIRPS and ARC2 do not show much variation but remain constant at around 30% rainy days. In contrast, CPC shows extensive variations, going from 20% in one year to 60% in another. TAMSAT tends to agree with CPC but lacks the wide swings. While ARC2 and CHIRPS report an average of 20% rainy days, and CPC and TAMSAT report 40%, MERRA-2 and ERA5, both reanalysis products, report rain 65% of days.

In Malawi, there is more agreement between ARC2/CHIRPS and CPC/TAMSAT, though again, CPC/TAMSAT tends to vary more. As with Ethiopia, MERRA-2 and ERA5 calculate a much higher share of rainy days.

In Niger, only MERRA-2 reports substantially more rainy days than others, though TAMSAT also reports slightly higher values.

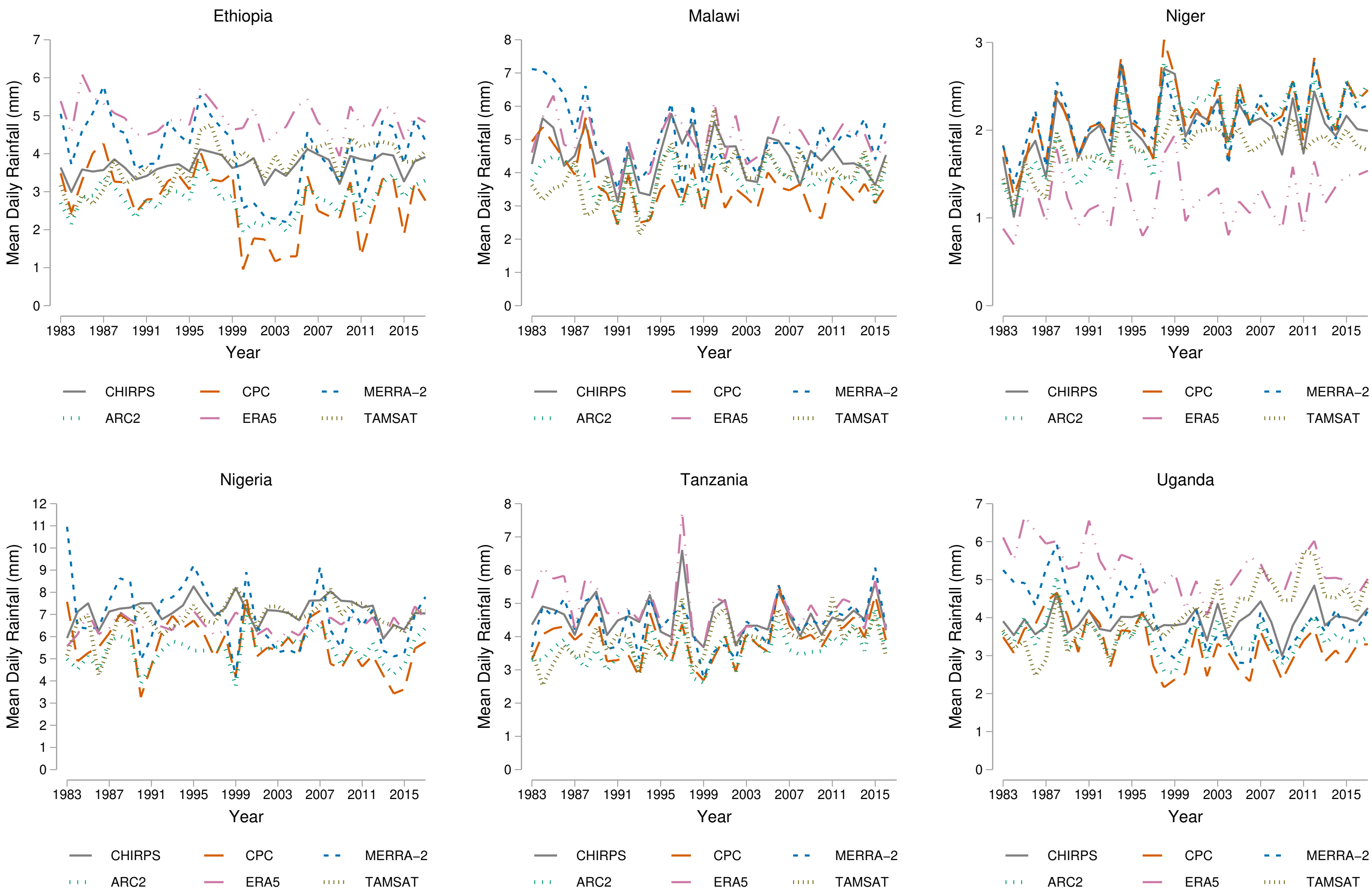
Nigeria repeats the pattern of a three-way separation. Here, CHIRPS, CPC, and ARC2 report the lowest share of rainy days, TAMSAT reports a slightly higher share of rainy days, and MERRA-2 and ERA5 report a substantially higher share of rainy days. According to ERA5, it rains in Nigeria 85% of the days, while MERRA-2 reports it rains 92% of the days.

Tanzania looks substantially like Nigeria, though TAMSAT is more in agreement with the non-reanalysis data products.

Similarly, in Uganda, the reanalysis data sets report rain nearly every day, while the other data sets are in substantial agreement. However, Uganda is unique in that it is the only country where TAMSAT reports fewer rainy days than ARC2/CHIRPS/CPC. In all other countries, TAMSAT tended to be between those data products and the reanalysis products.

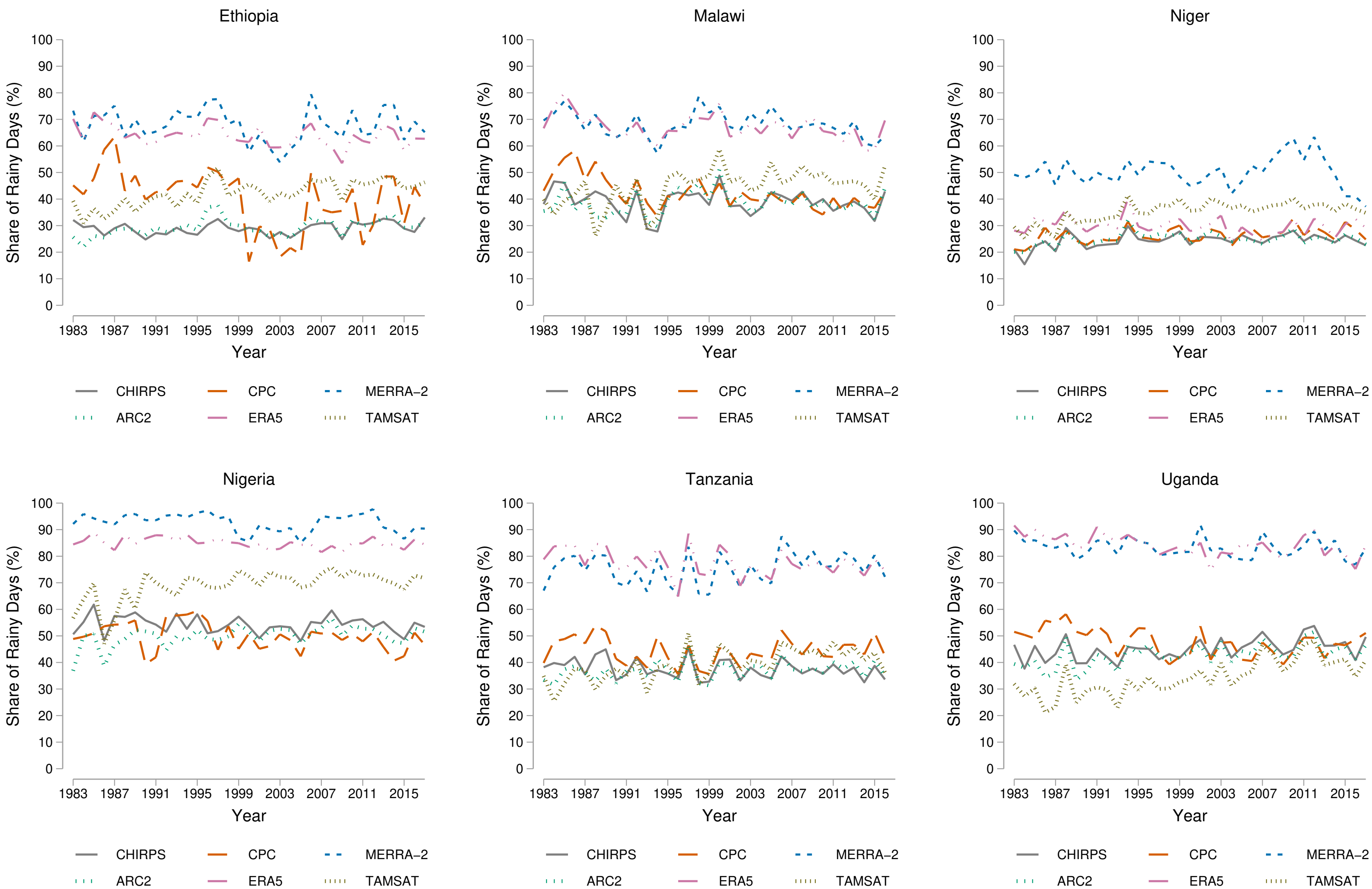
There are evident differences for any given rainfall metric across country and between remote sensing products. That differences exist across metrics and country is not surprising. What is more surprising are the differences across products for a single metric in a single country. When we examine the share of rainy days, these differences become stark. Our descriptive analysis highlights three important points. First, rainfall metrics calculated by different remote sensing products differ, and those differences vary by country. Second, for a single remote sensing product, there is often inter-country heterogeneity in how it performs. ERA5 tends to report the largest value for mean daily rainfall except in Niger where it reports the least. TAMSAT tends to report a larger share of rainy days than ARC2/CHIRPC/CPC, except in Uganda, where it reports the fewest. This highlights an important point: remote sensing products calculate the same metrics in a notably different manner depending upon the country. There is a geographical component to how accurate a product is that has not been explored or understood by the economics literature. Third, in the same country, while looking at different rainfall metrics, remote sensing products may show some relationships in one rainfall metric, whereas, in another rainfall metric, we see an entirely new grouping of products or no grouping at all. As an example, remote sensing products calculated from merged gauge, OI on gauge data, or reanalysis data sets show different patterns of results in the same country for different rainfall metrics, meaning there is not only inter-country heterogeneity but also intra-country heterogeneity.

Figure 4.1: Mean Daily Rainfall



Note: The figure presents the time series of the average value of the mean daily rainfall for each year in each country, measured by different remote sensing products. The y-axis has the mean daily rainfall in mm, and the x-axis has the duration of the analysis from 1983 to current.

Figure 4.2: Share of Rainy Days



Note: The figure presents the time series of the average value of the share of rainy days for each year in each country, measured by different remote sensing products. The y-axis has the share of rain days in percentages, and the x-axis has the duration of the analysis from 1983 to current.

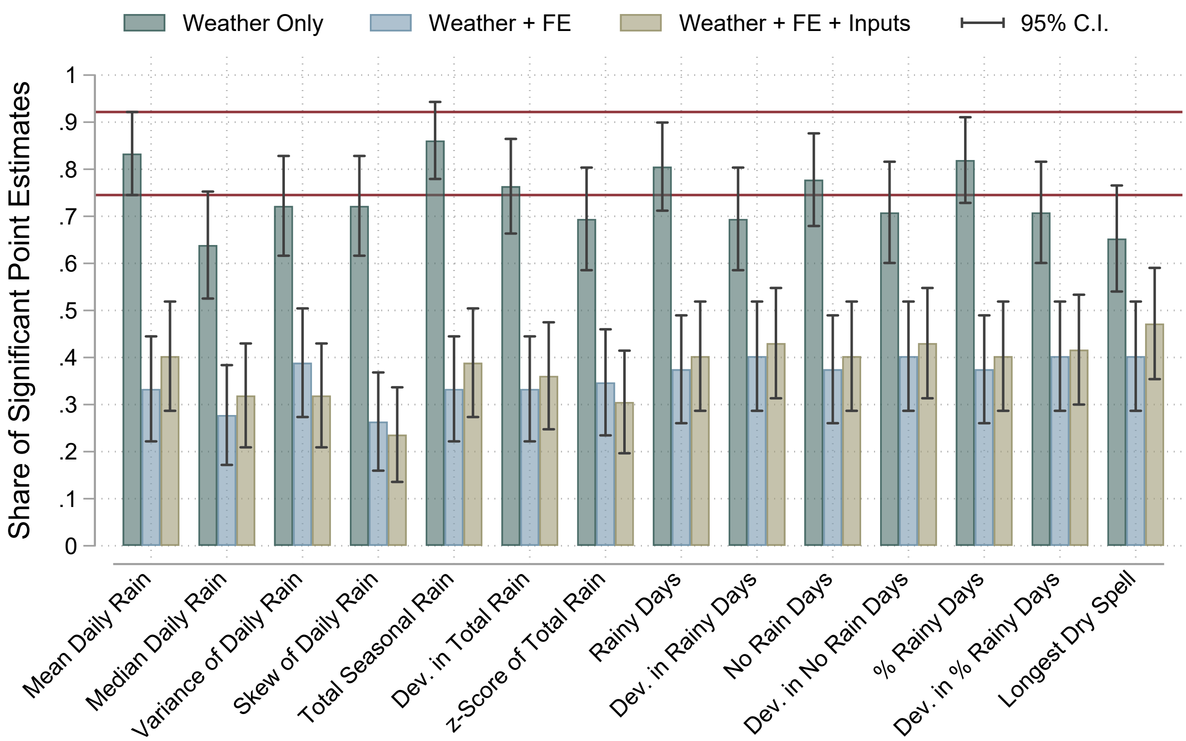
4.2 Share of Significant of p-values

Having discussed the inter and intra-country variation in rainfall metrics coming from different remote sensing products, we now shift to try and understand the predictive power of each rainfall metric. As we saw in Section 2.3, there is little agreement in the economics literature about which rainfall metric to use, even in the same context. To answer our question about the impact of using one rainfall metric over another, we estimate typical first stage IV regression of agricultural productivity on different rainfall metrics. Then we calculate the share of significance coefficients. A larger share of p-values less than 0.05, the stronger the metric is as an IV. The smaller the share, the weaker the instrument.

Figure 4.3 displays the share of coefficients for each rainfall metric, by each of our three econometric specifications: weather only, weather and household fixed effects, weather, household fixed effects, and measured inputs. These are aggregated over country, remote sensing product, and outcome variable. The figure summarizes the results of 3,024 regressions, with each column representing 72 regressions. For a column, which is a weather metric-specification pair, we calculate the mean number of regressions in which the coefficient on the weather metric is significant at $p < 0.05$. We also calculate the standard deviation of the mean. We divide the mean number of regressions with significant coefficients by 72 to calculate the share of significant point estimates. We use the standard deviation to calculate 95% confidence intervals around the mean. We can then apply our heuristic difference test to determine if one weather metric is a substantially stronger/weaker instrument than another metrics. We have also drawn horizontal lines to mark the upper and lower bounds of the confidence interval on mean daily rainfall so as to allow for easy visual inspection of where differences exist.

For regressions that only include weather, total rainfall has the highest share of significance, whereas median daily rainfall has the lowest share of significance. Many weather metrics are significant between 70% and 80% of the time. The exceptions are median, variance, skew, z-score, deviations in rainy days, deviations in no rain days, deviations in percent rainy days, and longest dry spell. All are weak instruments in terms of weak difference heuristics when compared to the mean daily and total seasonal rainfall. When we compare rainfall metrics

Figure 4.3: Share of Significance of Point Estimates



Note: The figure displays the share of significant coefficients for each rainfall metric, by each of our three econometric specifications, aggregated over country, remote sensing product, and outcome variable. The figure summarizes the results of 3,024 regressions, with each column representing 72 regressions. For a column, we calculate the mean number of regressions in which the coefficient on the weather metric is significant at $p < 0.05$. We also calculate 95% confidence intervals around the mean. Horizontal lines mark the upper and lower bounds of the confidence interval on mean daily rainfall.

to our strongest IV (total seasonal rainfall), we find that the median and longest dry spells fail our strong difference test, suggesting they are particularly weak options among these fourteen potential IVs.

Looking at the share of significant point estimates on weather metrics in the weather only specification, there is a clear hierarchy of potentially stronger and weaker rainfall instruments. But, including only weather as an explanatory variable ignores the fact that we have panel data. Those regressions will only be correctly specified if unobserved time-invariant household heterogeneity, currently in the error term, is uncorrelated with weather. If that is the case, then weather can serve as an instrument for agricultural productivity and total seasonal rainfall appears to be a particularly strong IV. If, however, weather is external to time-invariant household unobservables but not exogenous to those unobservable, our weather only specification will be mis-specified and rainfall may not pass the exclusion restriction.

To test this hypothesis, we control for household fixed effects and examine how rainfall changes as predictor of agricultural productivity. As Figure 4.3 demonstrates, most weather metrics go from being significant 80% of the time to significant only about 35% of the time. Adding measured inputs to the fixed effects does not substantially alter the estimates. Once we use fixed effects to control for time-invariant household unobservables, every metric performs similarly, with no difference in metrics based on either our strong or weak difference test.

To explain this, we conjecture that while rainfall is external to household characteristics, it is not exogenous in that rainfall is correlated with time-invariant household characteristics and if those are left uncontrolled rainfall is not orthogonal to the error term. This means that our weather only specification, where rainfall metrics appeared to be strong IVs, is mis-specified, with coefficients on rainfall metrics being biased upward. The most obvious candidate household characteristic that is correlated with rainfall is geographic location. There are time-invariant characteristics about the geographic location that determine how much rain the location gets but that also, independent of rainfall, are correlated with agricultural productivity. Absent a location fixed effect, rainfall is unlikely to satisfy the exclusion restriction.

This potential exclusion restriction violation has an easy solution - include control for household or location fixed effects. The problem with this solution, at least in the context of a first stage regression in which rainfall predicts agricultural productivity, is that controlling for household fixed effects creates a potential weak instrument problem. In 65% of fixed effects regressions, rainfall is not a significant predictor of agricultural productivity, meaning it frequently no longer satisfies the relevancy assumption. There is a clear trade-off, in our setting, between satisfying the exogeneity assumption and the relevancy assumption. We explore this more in the next section where we examine coefficients on specific regressions and attempt to draw conclusions regarding which rainfall metrics might remain good IVs and in what context.

4.3 Coefficients

In this section, we determine which metrics might serve as suitable instruments based on the consistency of their sign and the significance of their coefficients. We know that rainfall metrics calculated from different remote sensing products differ and that those differences vary across countries. We present a series of specification charts by country, which allows us to examine the coefficient size and significance by metric, remote sensing products, and dependent variable. We limit our analysis to just specifications with rainfall, household fixed effects, and measured inputs.

We have 168 regressions in every figure. The "$\diamond$" represents the coefficient of rainfall metric and it comes from a combination of rainfall metric, remote sensing product, and the outcome variable. The figure is divided into two sections: on the left of the red line are coefficients less than zero, whereas on the right of the red line are those greater than zero. We also have symbols for the significance of the coefficient: "+" when the negative or positive coefficients are significant at 0.05 or "Δ" when the negative or positive coefficients are not significant at 0.05. We have our fourteen rainfall variables, the six remote sensing products, and the two outcome variables on the y-axis. Our x-axis represents each individual regression.

To analyze the specification charts, we first look for groupings of different rainfall metrics. This allows us to gain a sense of which rainfall metrics might be good substitutes for each other and sheds light on the appropriateness of using many metrics to predict the same outcome, which we saw evidence of in Figure 1.1. Second, we determine which rainfall metrics are statistically significant (relevant). If they are significant, we consider their sign; if they are not, we discard them. Third, for rainfall metrics with predominantly significant coefficients, we look at the number of coefficients that are negative and positive. A good instrument would be the one that has a vast majority of coefficients of the same sign (consistent). Finally, we examine if the coefficients come from a particular remote sensing product and if the sign and significance depends on the country. This allows us to give recommendations and guidance both about which rainfall variables are potential good instruments for a broad context and which are only applicable in a narrow set of circumstances.

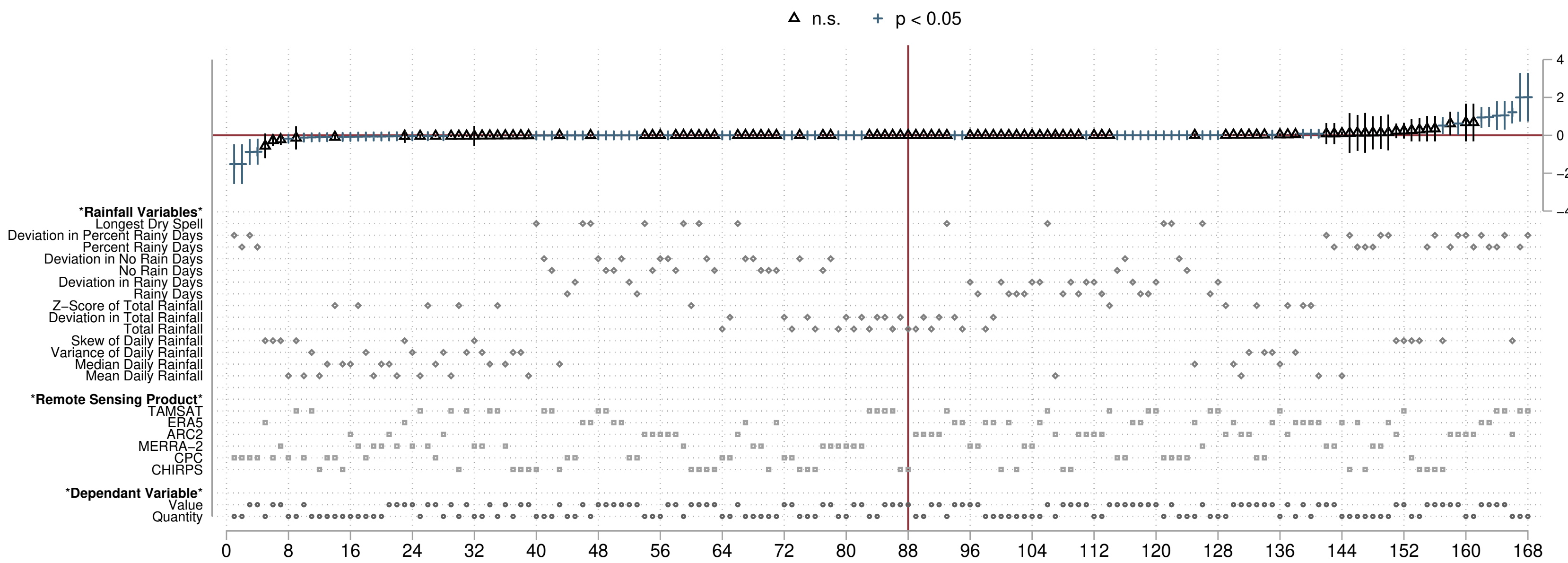
4.3.1 Ethiopia

Looking at the pattern of coefficient grouping in Figure 4.4, we can see that mean, median, and variance have coefficients of similar sign and size. Similarly, total rainfall and its deviation form a group centered around zero. Rainy days (and its deviation) and the percentage of rainy days (and its deviation) form two groups of similarly sized positive coefficients. Not surprisingly, the number of days without rain (and its deviation) mirror rainy days by forming a group of similarly sized negative coefficients. The remaining variables (skew, z-score, longest dry spell) fit into no obvious group because the size and sign of the coefficients on these metrics are widely dispersed from large and negative to large and positive. We conclude that, at least for Ethiopia, mean, median, and variance could be used as substitute IVs, while rainy days, no rain days, and percent rainy days (and their deviations) could be used as substitutes with the understanding that no rain days would give the opposite sign of the others.

We next turn to examine the number of significant coefficients for the rainfall metric. In Ethiopia, seventy-two coefficients are significant (43%). Each rainfall metric appears in twelve regressions and, in Ethiopia, all have around 40%-50% of coefficients that are significant. The lone exception is skew, with only two of twelve significant coefficients (17%). This suggests that, except for skew, no rainfall metric is particularly more relevant (stronger IV) than any other metric in Ethiopia.

As we cannot distinguish between a potentially good IV and a potentially bad IV based on relevance, we next look at the sign on the significant coefficients for each metric. This helps us understand what rainfall metrics may give mixed or inconsistent results when predicting agricultural productivity. The significant coefficients for the grouping of mean, median, and variance are nearly all negative, suggesting they might make good instruments. Median, in particular, is always negative and is the only variable among the 14 that always has a consistent sign. For mean and variance, there is an exception to the consistency of their sign. If the data for calculating these metrics come from ERA5, the sign switches so that their relationship with agricultural productivity becomes positive. Similar to mean/median/variance, total rainfall and deviations in total are relevant and consistent, making them good IV candidates. Each are significant in half of their regressions and the coefficients are negative in all but one case: when the data comes from ERA5. Among the remaining variables with five or six of twelve significant coefficients (rainy days, no rain days, percent rainy days, and their deviations), none have a consistent relationship with agricultural productivity. For each of these variables, about half of their significant coefficients express a positive relationship with agricultural productivity, and half express a negative relationship. The positive coefficients on these metrics tend to come from TAMSAT, while the negative coefficients tend to come from ARC2, but these relationships are not absolute nor without their exceptions.

Figure 4.4: Ethiopia



Note: The figure presents a specification curve of 168 regressions. It is divided into two sections: on the left of the red line are those coefficients less than zero, whereas on the right of the red line are those greater than zero. Significant and non-significant coefficients are designated above, where + signifies that the negative or positive coefficients are significant at 0.05, whereas Δ signifies that coefficients are not significant. There are fourteen rainfall variables, six remote sensing products, and two outcome variables on the y-axis. Our x-axis represents each individual regression, sorted by coefficient size.

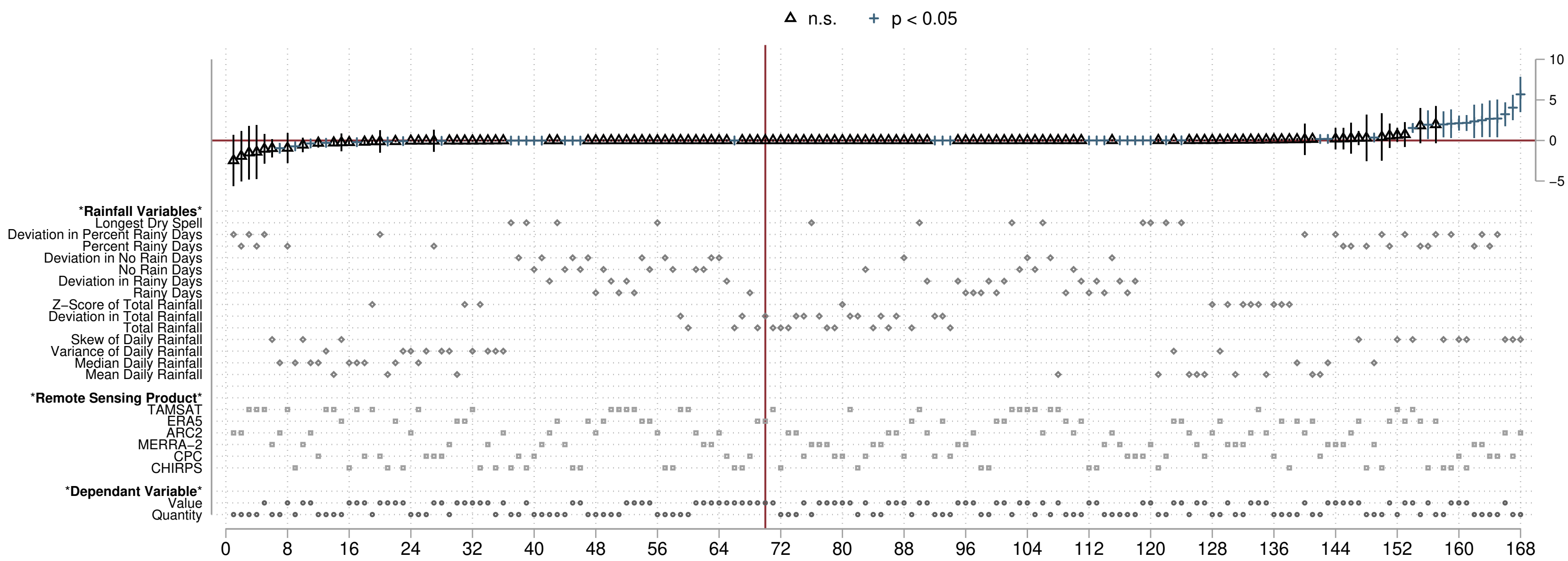
4.3.2 Malawi

Turning to Malawi (Figure 4.5), we see that median and variance have coefficients of similar size and sign. Unlike in Ethiopia, mean in Malawi does not provide similar predictions as median and variance. Instead, mean produces mostly positive coefficients. Total rainfall (and its deviation) are centered around zero and mostly positive. Rainy days (and its deviation), and the percentage of rainy days (and its deviation), like Ethiopia, are two groups of similarly sized positive coefficients. The number of days without rain (and its deviation) are a mirror of rainy days (and its deviation), forming a group of similarly sized negative coefficients. The remaining rainfall metrics (skew, z-score, and the longest dry spell) do not form groups with any other rainfall metrics, as they have different sizes and signs for their coefficients. We conclude that, in Malawi, apart from rainfall metrics with deviations, median and variance could be used as substitute IVs.

Malawi only has forty-six significant coefficients (27%) for the rainfall metrics, far fewer than Ethiopia. Most rainfall metrics have significant coefficients only 15%-25% of the time. The exceptions to this are median daily and the longest dry spell, both of which have six of twelve significant coefficients (50%), and skew of daily rainfall, which has seven of twelve significant coefficients (58%). The z-score of total rainfall has no significant coefficients in Malawi that could predict agricultural productivity.

To distinguish between a suitable and a weak IV, we look at the sign on the significant coefficients of rainfall metrics. Mean has an inconsistent relationship with agricultural productivity. The six significant coefficients for median are all negative, and thus, it might make a good instrument. The exception is if median is calculated from MERRA-2 data, which results in the coefficients switching sign from negative to positive. While variance forms a group with median, it performs worse in that only three of the twelve coefficients are significant. Skew of daily rainfall is the most relevant predictor of agricultural productivity in Malawi, with 58% of the coefficients being significant, as well as being among the most consistent, with all being negative. Longest dry spell also performs well, as it has mostly positive significant coefficients unless data comes from CHIRPS, which switches the coefficient sign from positive to negative. Coefficients of all of the remaining metrics are rarely significant, making them weak instruments in Malawi.

Figure 4.5: Malawi



Note: The figure presents a specification curve of 168 regressions. It is divided into two sections: on the left of the red line are those coefficients less than zero, whereas on the right of the red line are those greater than zero. Significant and non-significant coefficients are designated above, where + signifies that the negative or positive coefficients are significant at 0.05, whereas Δ signifies that coefficients are not significant. There are fourteen rainfall variables, six remote sensing products, and two outcome variables on the y-axis. Our x-axis represents each individual regression, sorted by coefficient size.

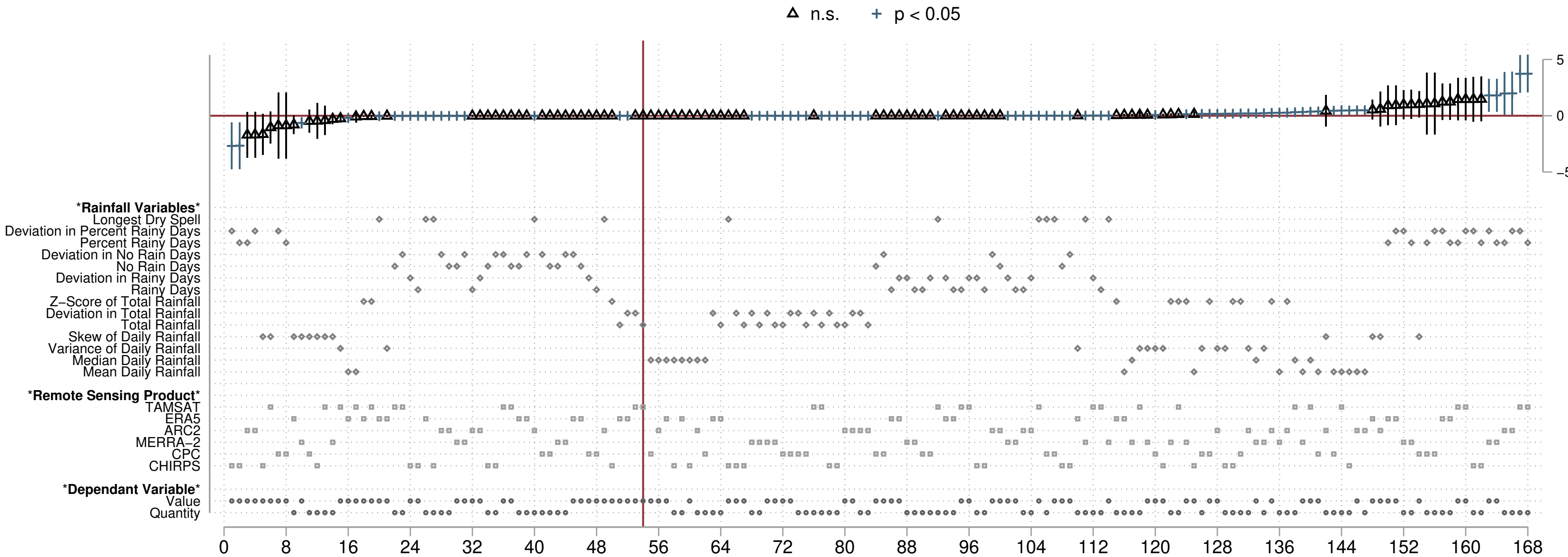
4.3.3 Niger

Looking at the pattern of coefficient grouping in Figure 4.6, we can see that mean and variance have mostly positive coefficients with similar sizes. Unlike in Ethiopia and Malawi, median is not a part of a group. Most of its coefficients are precisely estimated zeros as the median value of rain in Niger is zero. Other groupings of similarly sized positive coefficients include total rainfall (and its deviation), rainy days (and its deviation), and percent rainy days (and its deviation). The number of days without rain (and its deviation) gives a similar size but different sign coefficients as the number of days with rainfall (and its deviation). The remaining variables (skew, z-score, and the longest dry spell) are not part of any group as the size and the sign of the coefficients on these metrics vary from positive to negative. We conclude that in Niger, mean and variance, and other rainfall metrics with their deviations, could be used as substitute instruments.

For the rainfall metrics in Niger, seventy-six coefficients are significant (45%). This value hides what is a nearly bimodal distribution, with six of the fourteen rainfall metrics being significant at least 50% of the time while the remainder are significant less than 30% of the time. Among the candidates for a strong IV are mean (9 of 12 significant), median (6 of 12), total (9 of 12) with its deviation (8 of 12), z-score (6 of 12), and the longest dry spell (10 of 12). This means in Niger, unlike Ethiopia and Malawi, we have a large pool of potentially strong IVs.

For a candidate metric to be a suitable IV, it must be both relevant and consistent. The significant coefficients of the grouping of mean and variance in Niger are nearly all positive, suggesting they might make good instruments. The lone exception is if mean is calculated using ERA5, which switches sign to negative. Total rainfall, along with its deviation and z-score, are relevant and consistent, making them good IV candidates. The exception to this is if the data come from ERA5, which again switches the sign on significant coefficients from positive to negative. The longest dry spell in Niger is the most relevant predictor of agricultural productivity in any country and for any metric. Its coefficients are significant in ten out of twelve regressions. However, it is an inconsistent predictor. Six of the ten coefficients are positive, suggesting that the longest dry spell increases output. Only in four regressions is the longest dry spell negatively correlated with output as expected. Despite this inconsistency of the longest dry spell, Niger has a stronger and more consistent set of candidate rainfall instruments than either Ethiopia or Malawi.

Figure 4.6: Niger



Note: The figure presents a specification curve of 168 regressions. It is divided into two sections: on the left of the red line are those coefficients less than zero, whereas on the right of the red line are those greater than zero. Significant and non-significant coefficients are designated above, where + signifies that the negative or positive coefficients are significant at 0.05, whereas Δ signifies that coefficients are not significant. There are fourteen rainfall variables, six remote sensing products, and two outcome variables on the y-axis. Our x-axis represents each individual regression, sorted by coefficient size.

4.3.4 Nigeria

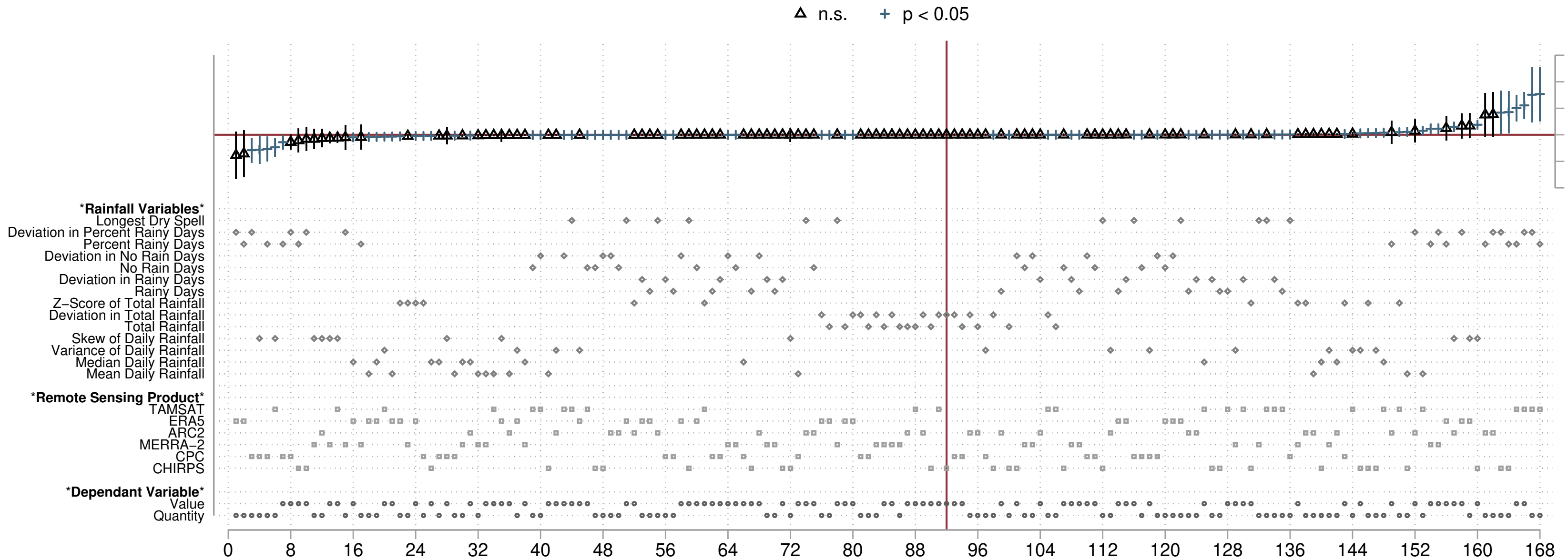
Figure 4.7 presents the specification curve for Nigeria, where we can see that mean and median have coefficients of similar sign and sign. This breaks from the patterns in Ethiopia, Malawi, and Niger, where we saw median and variance always paired. Total rainfall and deviation in total rainfall form a group centered around zero. Unique to Nigeria, rainy days (and their deviations), days without rain (and their deviation), and the share of rainy days (and their deviation) all form groups. These groupings are divided into two clusters, one of negative coefficients and one of positive coefficients. The remaining variables (variance, skew, z-score, and the longest dry spell) conform to no group and have signs and sizes widely from large and positive to large and negative. We conclude that in Nigeria, mean and median could be used as substitute instruments, as well as total and its deviation. For the other rainfall variables, the groupings appear split between showing positive and negative relationship with agricultural productivity.

Of the regressions in Nigeria, seventy-one coefficients are significant (42%). Each rainfall metric appears in twelve regressions, and like Ethiopia, all perform about the same (40%-50% significant coefficients). This differs from Malawi, which had almost no strong predictors of agricultural productivity, and Niger, which had a number of metrics highly correlated with agricultural productivity. We conclude that in Nigeria, like Ethiopia, no metric is particularly more relevant than another.

Turning to the sign of the significant coefficients, we try to determine which rainfall metrics are consistent predictors of our outcome. Of the six significant coefficients on median five are negative. The sole exception is if data comes from ARC2, which results in median switching signs from negative to positive. Rainy days (and its deviation) and no rain days (and its deviation) form two groups which have five of six of their significant coefficients on one side of zero (positive for rainy days and negative for no rain days). The exception in all four metrics is when data come from CPC, which switches the dominant signs. The remaining rainfall metrics do not have a consistent relationship with agricultural productivity. Half the coefficients of these metrics show a positive relationship with outcomes, and half show a negative relationship. While Nigeria has about the same number of relevant metrics as

Ethiopia, there is much less consistency in the sign on these metrics. This means that there are few rainfall metrics that are good (relevant and consistent) potential IVs.

Figure 4.7: Nigeria



Note: The figure presents a specification curve of 168 regressions. It is divided into two sections: on the left of the red line are those coefficients less than zero, whereas on the right of the red line are those greater than zero. Significant and non-significant coefficients are designated above, where + signifies that the negative or positive coefficients are significant at 0.05, whereas Δ signifies that coefficients are not significant. There are fourteen rainfall variables, six remote sensing products, and two outcome variables on the y-axis. Our x-axis represents each individual regression, sorted by coefficient size.

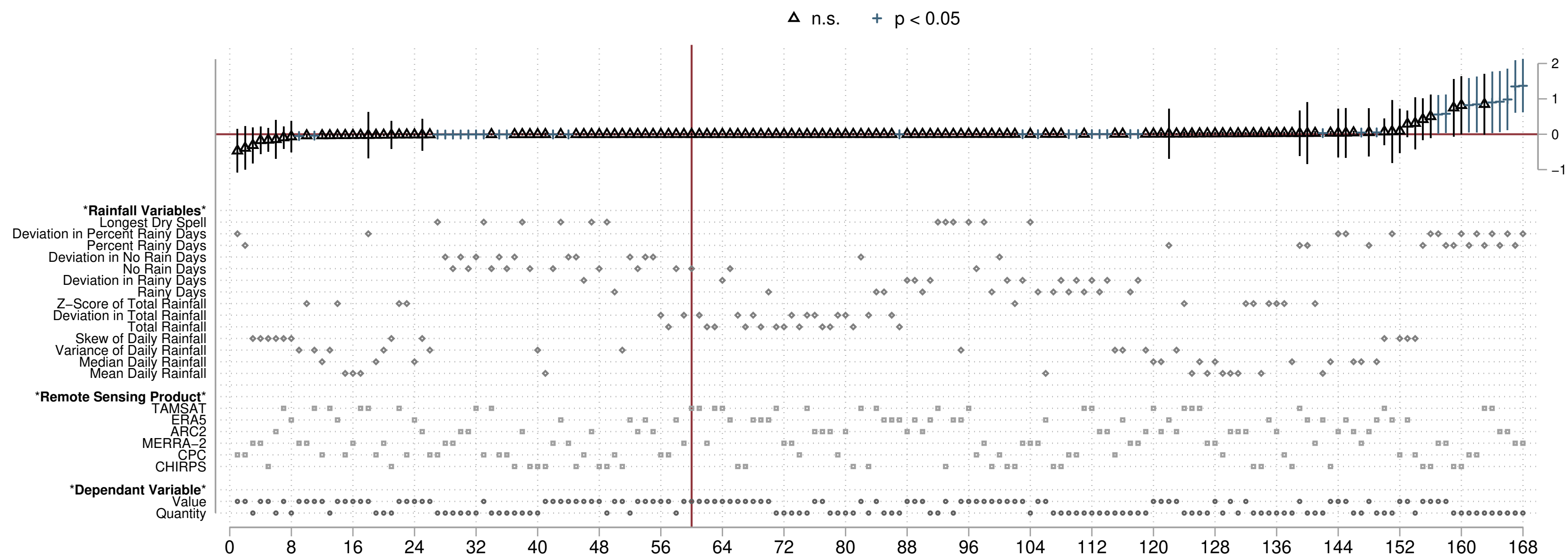
4.3.5 Tanzania

Turning to Tanzania (Figure 4.8), we see many of the same patterns as in other countries. Mean and median have coefficients of similar sign and size. Coefficients on total rainfall and its deviation are nearly all positive and grouped around zero. Other groupings include rainy days (and its deviation) with mostly positive coefficients and no rain days (and its deviation) with mostly negative coefficients, with all four having similar sizes. Percentage of rainy days (and its deviation) has mostly large and positive coefficients. As is the case in other countries, the remaining variables (variance, skew, z-score, and the longest dry spell) do not fit into any obvious groups because of their varied sizes and different signs. We conclude that in Tanzania, mean/median and other rainfall variables/their deviations would be appropriate substitutes as they would not change the relationship with agricultural productivity.

Of our six countries, Tanzania has the least number of significant coefficients for rainfall metrics at thirty-five (21%). Some rainfall metrics (skew, deviations in total rainfall, and the z-score) do not produce any significant coefficient. Half of the remaining rainfall metrics (mean, median, variance, total, and the longest dry spell) are significant around 10%-15% of the time, and the other half (rainfall days, no rain days, the percentage of rainy days, and their three deviations) are significant around 30%-40% of the time. No metric is significantly correlated with outcomes, even 50% of the time. Relative to rainfall in other countries, rainfall lacks relevance (weak IV) in Tanzania.

While there appear to be a few potentially suitable IVs in Tanzania, we still categorize metrics based on the consistency of their significant coefficients. In this respect, Tanzania, too, is an outlier. All the rainfall metrics with significant coefficients have consistent signs. That is, if the rainfall metric in Tanzania has significant coefficients, the signs are either always positive (mean, median, total, rainy days and its deviation, and percent rainy day and its deviation) or always negative (variance, no rain days, and its deviation, and the longest dry spell). There is no rainfall metric that produces both significantly positive and significantly negative coefficients. While most rainfall metrics in Tanzania are not relevant, when a metric is relevant, it has a consistent relationship with agricultural productivity.

Figure 4.8: Tanzania



Note: The figure presents a specification curve of 168 regressions. It is divided into two sections: on the left of the red line are those coefficients less than zero, whereas on the right of the red line are those greater than zero. Significant and non-significant coefficients are designated above, where + signifies that the negative or positive coefficients are significant at 0.05, whereas Δ signifies that coefficients are not significant. There are fourteen rainfall variables, six remote sensing products, and two outcome variables on the y-axis. Our x-axis represents each individual regression, sorted by coefficient size.

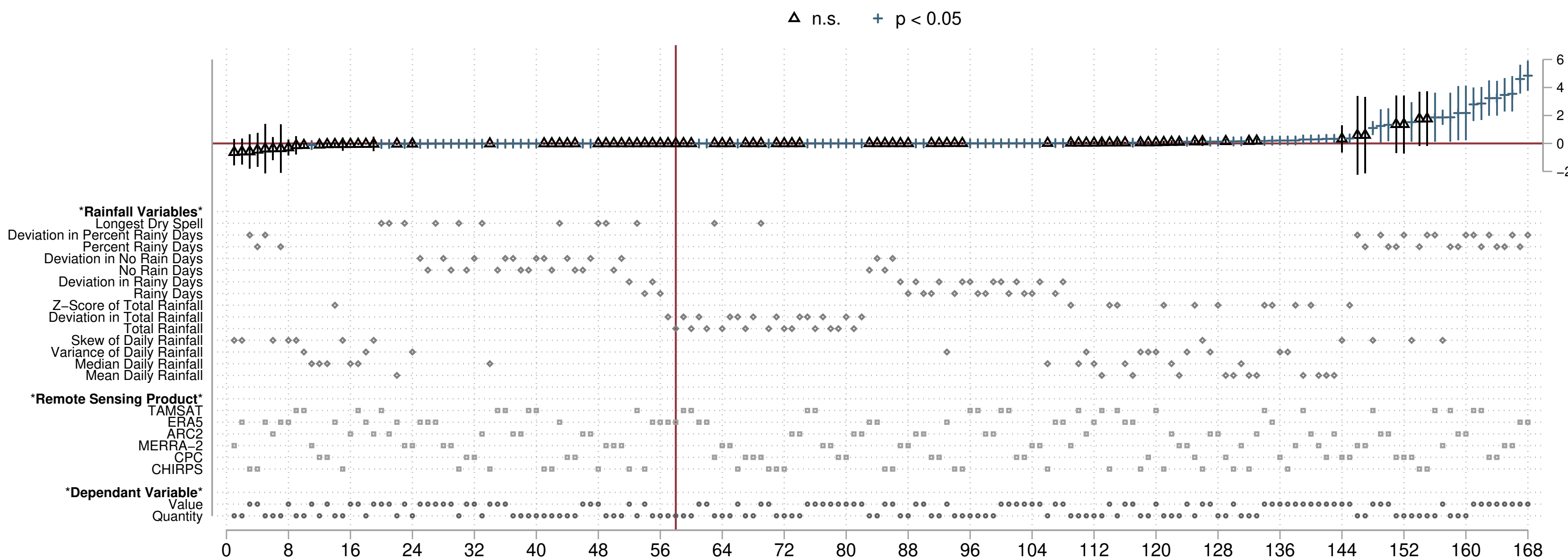
4.3.6 Uganda

Consider the pattern of coefficient grouping for Uganda presented in Figure 4.9. While Tanzania has the fewest relevant rainfall metrics, Uganda has the most. However, like Tanzania, rainfall metrics in Uganda are consistent predictors of agricultural productivity. Uganda, unlike other countries, does not have any groupings between mean, median, and variance. Mean has nearly all coefficients greater than zero, median has six negative and six positives, and variance has three negative and nine positive. As with most other countries, total rainfall and its deviation are clustered around zero. Rainy day (and its deviation) and percent rainy days (and its deviation) form two groups of similarly sized positive coefficients. Whereas no rain days and its deviations form a group of similarly sized negative coefficients. The remaining rainfall metrics (skew, z-score, and the longest dry spell) form no groups with other rainfall variables in terms of sign and size. We conclude that apart from rainfall metrics with their deviations, no other rainfall metrics can be used as a substitute that would produce coefficients with consistent sign and size.

Uganda has the highest number of significant coefficients among the countries we analyze, with eighty-one (48%). Eleven of the fourteen rainfall metrics have coefficients that are significant 50%-60% of the time. The exceptions are median, variance, and skew. Relative to other countries, rainfall appears to be a relevant (strong) instrument in predicting agricultural production.

In Uganda, rainfall metrics tend to be not just relevant but consistent. As in Tanzania, the rainfall metrics with significant coefficients have a consistent sign. Significant coefficients of every metric, with the exception of the median, always carry the same sign. If a metric has one positive and significant coefficient, then all significant coefficients are positive (mean, variance, skew, total rainfall and its deviation, rainy days and its deviation, and percent rainy day and its deviation). Similarly, if a metric has one negative and significant coefficient, then all are negative (no rain days and its deviation, and the longest dry spell). Compared to Tanzania and all other countries in the analysis, rainfall in Uganda is the most relevant and consistent predictor of agricultural productivity, suggesting that many rainfall metrics are good IV candidates in this setting.

Figure 4.9: Uganda



Note: The figure presents a specification curve of 168 regressions. It is divided into two sections: on the left of the red line are those coefficients less than zero, whereas on the right of the red line are those greater than zero. Significant and non-significant coefficients are designated above, where + signifies that the negative or positive coefficients are significant at 0.05, whereas Δ signifies that coefficients are not significant. There are fourteen rainfall variables, six remote sensing products, and two outcome variables on the y-axis. Our x-axis represents each individual regression, sorted by coefficient size.

4.4 Discussion

Having examined the results of 3,024 first-stage IV regressions across six countries, we draw conclusions and make recommendations for best practices. These relate to our three headline findings: one, there is substantial heterogeneity in the performance of rainfall metrics in predicting agricultural productivity; two, this heterogeneity is often a function of remote sensing sources; and three, most rainfall metrics are weak instruments once we control for household unobservables.

The following are our three conclusions and related recommendations. First, we find evidence that not all rainfall metrics are equally strong instruments, as not all rainfall metrics are significant predictors of (relevant to) agricultural productivity. Whether a rainfall metric is relevant often depends upon the remote sensing product the data comes from and the country it is used in. In other words, there is a large amount of heterogeneity in the way rainfall metrics perform across remote sensing products and countries. In Ethiopia, mean, total, deviation in total, and deviations in rainy days, no rainy days, and percent rainy days are good candidate instruments, as they are all relevant at least 50% of the time. The exceptions arise when some metrics (mean, total, and deviations in total) are calculated using ARC2 or TAMSAT. Another exception is when the deviation in metrics comes from CHIRPS or MERRA-2. What this means is that using data from these products in Ethiopia can result in inconsistent and unpredictable results. In Malawi, median, skew, and longest dry spell are good potential instruments. There are some exceptions to this, but these exceptions do not appear to be a function of selecting a particular remote sensing product. In Niger, mean, variance, total rainfall, deviation in total rainfall, the z-score of total rainfall, and longest dry spell are all strong and consistent predictors of agricultural productivity. The exception is when variance, deviation in total rainfall, and the z-score come from ERA5 or TAMSAT data. In those cases, these metrics are no longer relevant. In Nigeria, z-score, rainy days (and its deviation), no rain days (and its deviation), and percent rainy days are all good candidates for an IV. The exceptions arise when metrics are calculated using data from ARC2 and ERA5. In Tanzania, no rainfall metric is relevant even 50% of the time. However, in the rare case that a coefficient is significant, it has a consistent relationship to outcomes. Finally, in Uganda, a large number of metrics are relevant to predicting agricultural productivity. As with Malawi and Tanzania, relevant metrics appear not to be a function of the source of the precipitation data.

Second, there exist few rainfall metrics that are consistent predictors of agricultural productivity. In Ethiopia, mean, total rainfall, and deviations in total are consistent predictors and could be used as substitutes for each other. The exception to this consistency is if the data comes from ERA5, in which case the signs of all three metrics switches from positive to negative. In Malawi, median, skew, and longest dry spell are consistent, though median is negatively correlated with outcomes while the other two metrics are positively correlated. Here, using MERRA-2 data to calculate median or CHIRPS to calculate longest dry spell switches the sign of each metric. In Niger, mean, variance, total rainfall, and deviations in total rainfall, z-score are consistent predictors in that they are all positively correlated with outcome, making them good substitutes for one another. The exceptions are few, and all arise when ERA5 data is used to calculate a metric. In Nigeria, only rainy days (and its deviations) and no rain days (and its deviations) are consistent, though they have opposite signs. However, when CPC data is used, rainy days become negatively correlated with agricultural productivity, while no rain days become positively correlated. In Tanzania, while no metric is relevant, all are consistent in that when a metric is significant, it always carries the same sign. Finally, in Uganda, all eleven relevant metrics are consistent. If a metric is positive/negative and significant, then it always is regardless of what remote sensing data source is used. Thus, there are heavy potential substitutes for rainfall IVs in Uganda. While rainfall metrics are relevant and consistent in Uganda, Uganda is not the norm. The implication of this is that even though a researcher may use a rainfall metric in one country that is relevant and consistent, using that metric in another country, or even calculating it from a different remote sensing source, can produce a weak instrument or flip the relationship between that instrument with agricultural productivity.

Third, the inclusion of household fixed effects has a large impact on the relevance of weather metrics in the first stage IV regression. While all our takeaways speak to the relevancy assumption, one of the four major identification assumptions for an instrumental variable, our last takeaway also has implications for the exogeneity assumption. The rainfall metrics are significant 80% of the time without household fixed effects. However, with the introduction of household fixed effects, rainfall as a significant predictor of agricultural productivity falls to 35%. Clearly, including household fixed effects creates a weak instrument problem. But not including them means that the time-invariant unobservable household characteristics are part of the error term violating the exogeneity assumption. The most likely unobservable household characteristic that correlates with weather is the household's geographic location.

Combined, our three takeaways place the user of rainfall IVs in an unenviable position. Controlling for household fixed effects or location fixed effects satisfies the exogeneity assumptions of the IV while likely violating the relevancy assumption. Outside of abandoning the use of rainfall as an IV altogether, users of rainfall IVs will need to provide strong evidence that their chosen IV is among the small set of metrics that we have found both relevant and consistent when controlling for household (location) fixed effect. Of particular import will be to demonstrate the robustness of results to different data sources to ensure results do not reflect, at best, spurious correlation or, at worst, p-hacking.

CHAPTER 5

Conclusion

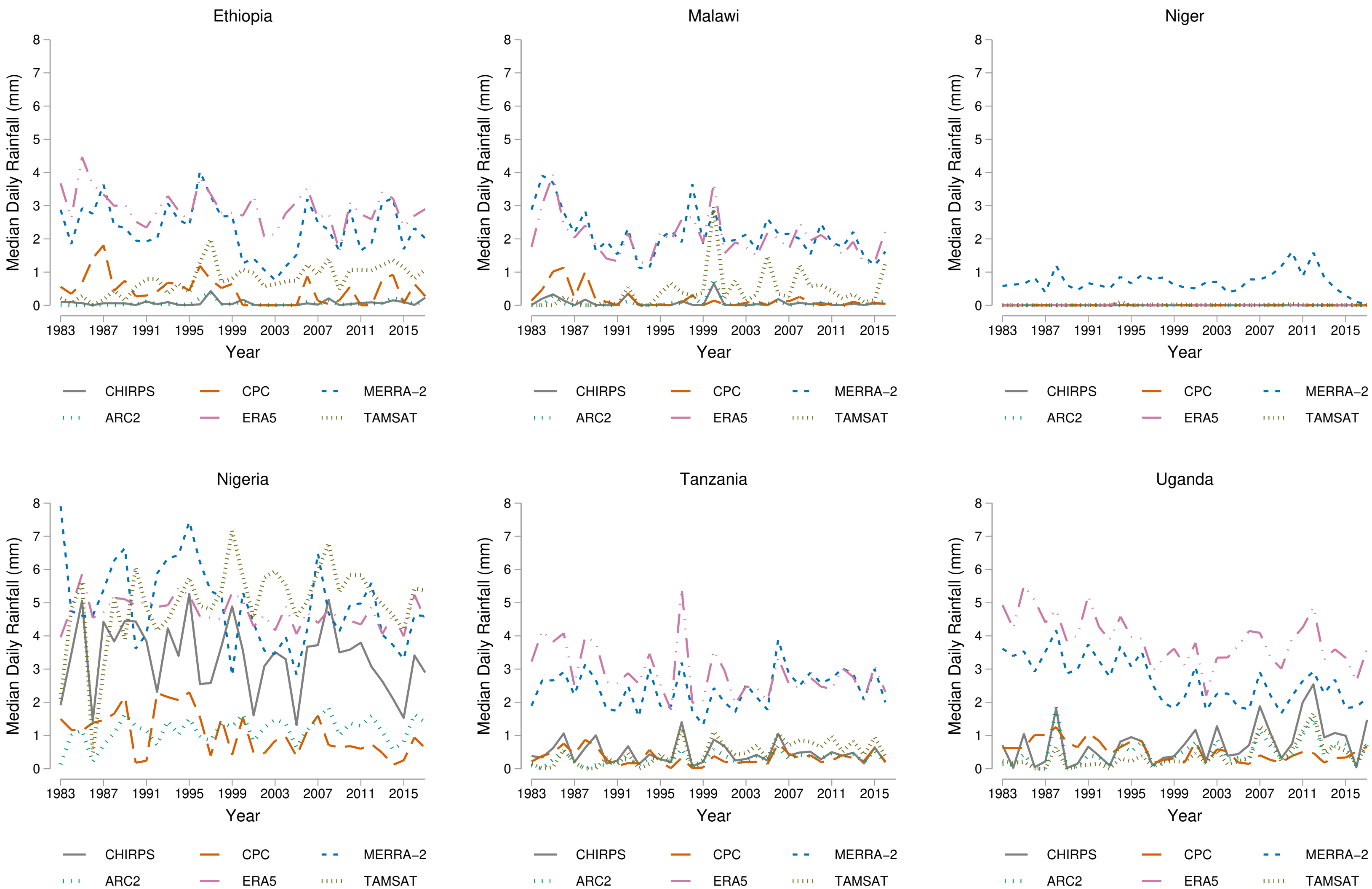
Given the low and declining cost of using remote sensing data, weather, in various measurements, has become a commonly used instrumental variable in economics. In this thesis, we have reviewed 174 papers and documented the ad-hoc nature of IV use in the field. To further examine the effects, we investigate the appropriateness of this ad-hoc approach, by combining six remote sensing weather products with the LSMS-ISA data from six countries. We looked for three elements in order to assess weather as an IV, specifically focusing on rainfall. First, we look for rainfall metrics that are consistent across remote sensing products and countries; next, we consider weather metrics that show a consistent relationship with agricultural productivity; and finally we seek to understand a degree of substitutability, weather metrics that have similar signs and significance to other weather metrics. Further, we considered the assumptions for IV use, focusing on an examination of the exogeneity of rainfall.

In this thesis, we found three things. First, we determined that a large amount of heterogeneity exists in the manner rainfall metrics perform across remote sensing products and countries. Next, we found that few metrics act as substitutes. We find that there is inconsistency in rainfall metrics in predicting agricultural productivity. And finally, we find that when controlling for household fixed effects, we encounter a weak instrument problem. Taken together, these findings contribute to the body of empirical literature in economics by examining the validity and exogeneity of instrumental variables, specifically weather and rainfall as an instrument. By providing guidance regarding which rainfall metrics could be used in lieu of another, depending upon the context. Our analysis confirms certain findings by other researchers about the existence of potential exclusion-restriction violations.

APPENDIX A

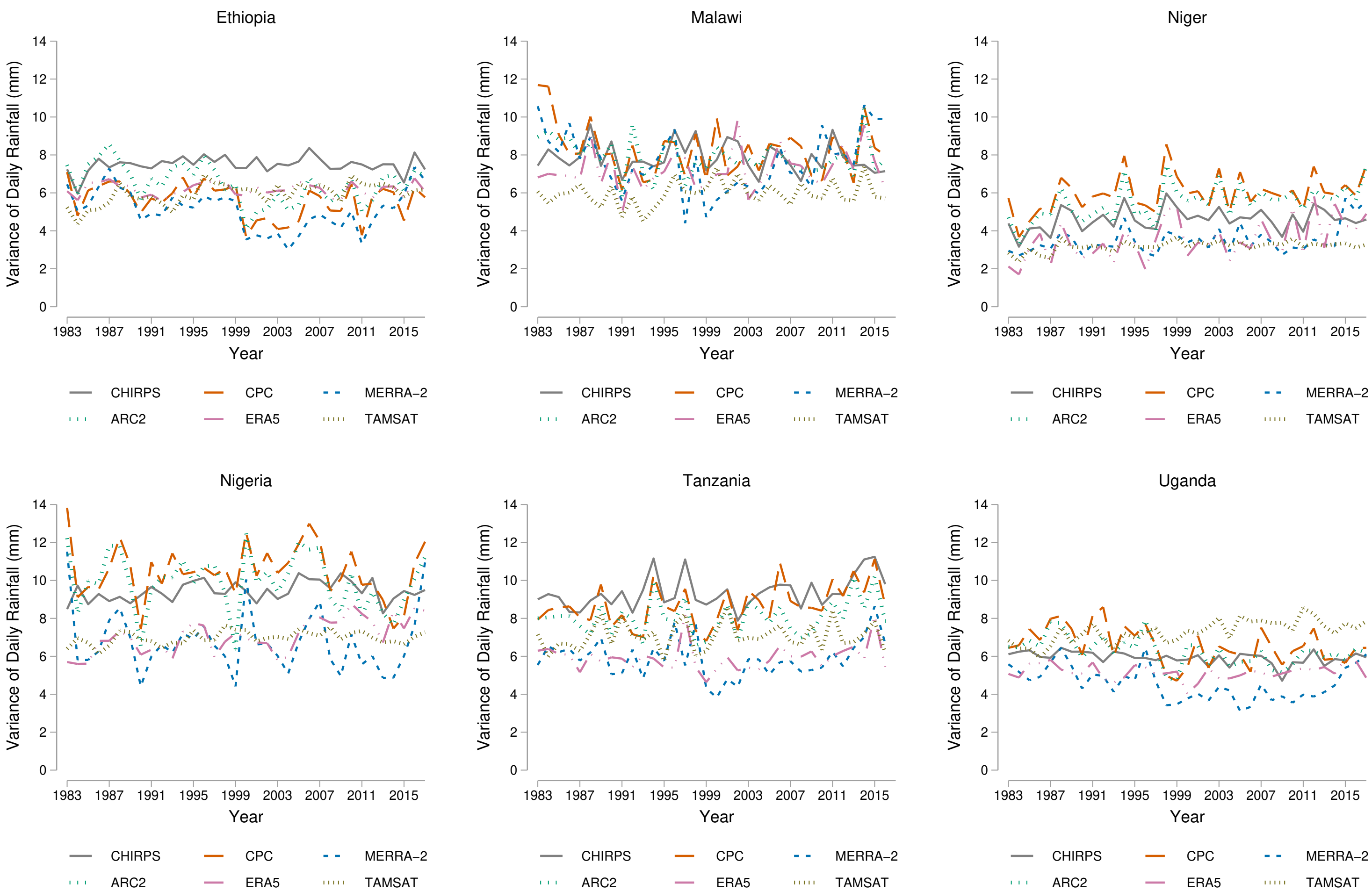
Descriptive Statistics

Figure A.1: Median Daily Rainfall



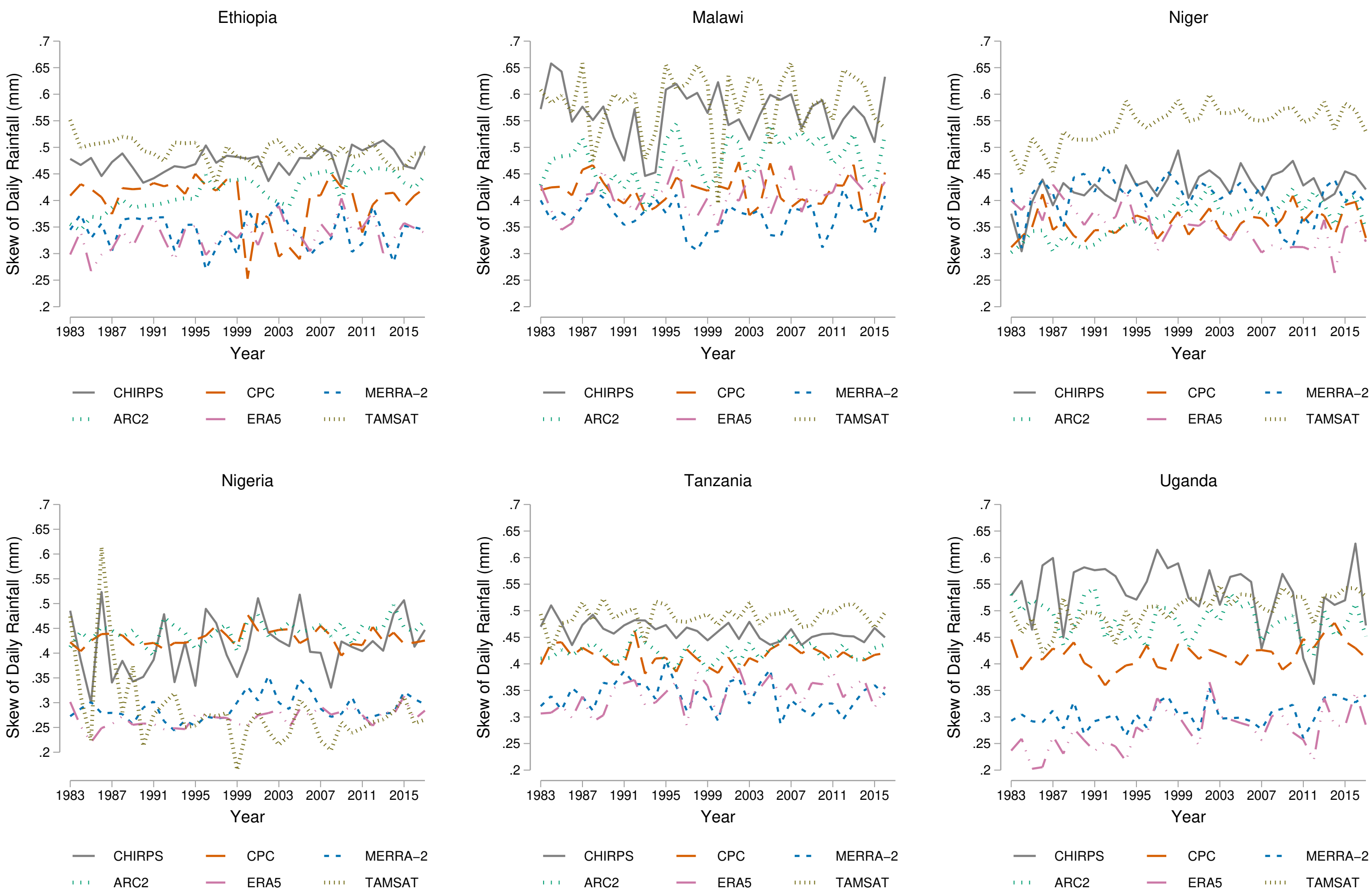
Note: The figure presents the time series of the average value of the median daily rainfall for each year in each country, measured by different remote sensing products. The y-axis has the median daily rainfall in mm, and the x-axis has the duration of the analysis from 1983 to current.

Figure A.2: Variance of Daily Rainfall



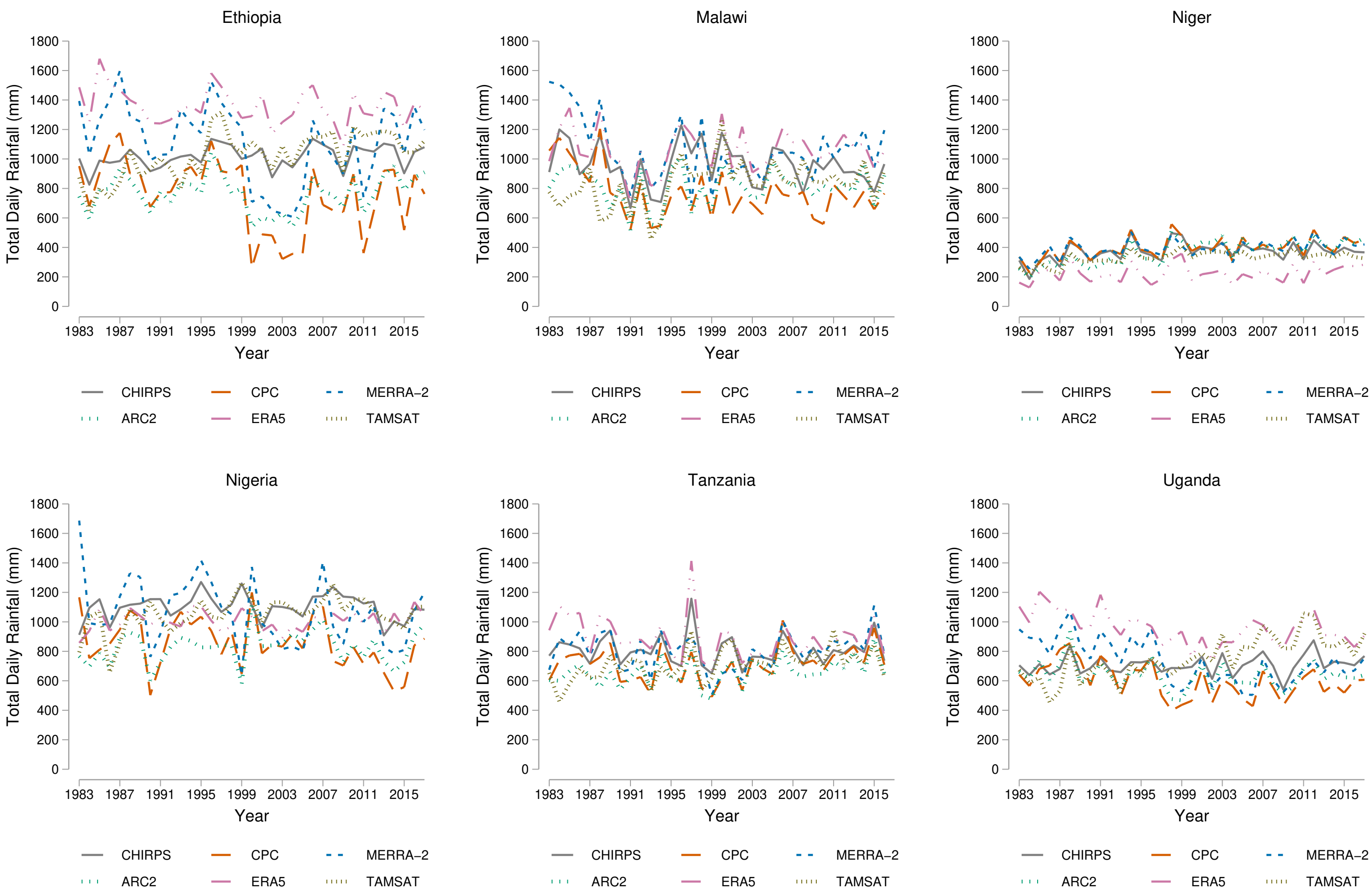
Note: The figure presents the time series of the average value of the variance of daily rainfall for each year in each country, measured by different remote sensing products. The y-axis has the variance of daily rainfall in mm, and the x-axis has the duration of the analysis from 1983 to current.

Figure A.3: Skew of Daily Rainfall



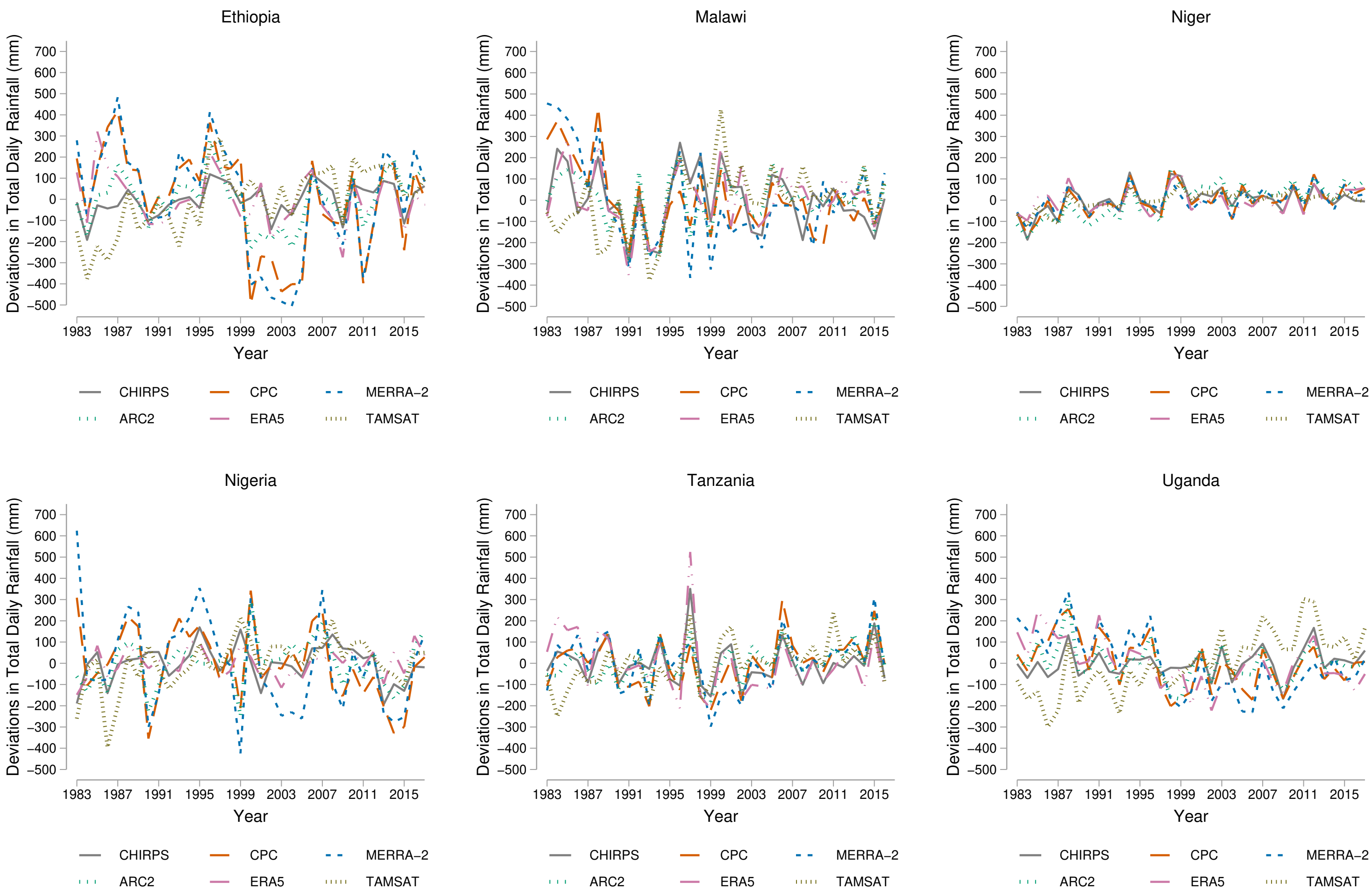
Note: The figure presents the time series of the average value of the skew of the daily rainfall for each year in each country, measured by different remote sensing products. The y-axis has the skew of daily rainfall in mm, and the x-axis has the duration of the analysis from 1983 to current.

Figure A.4: Total Daily Rainfall



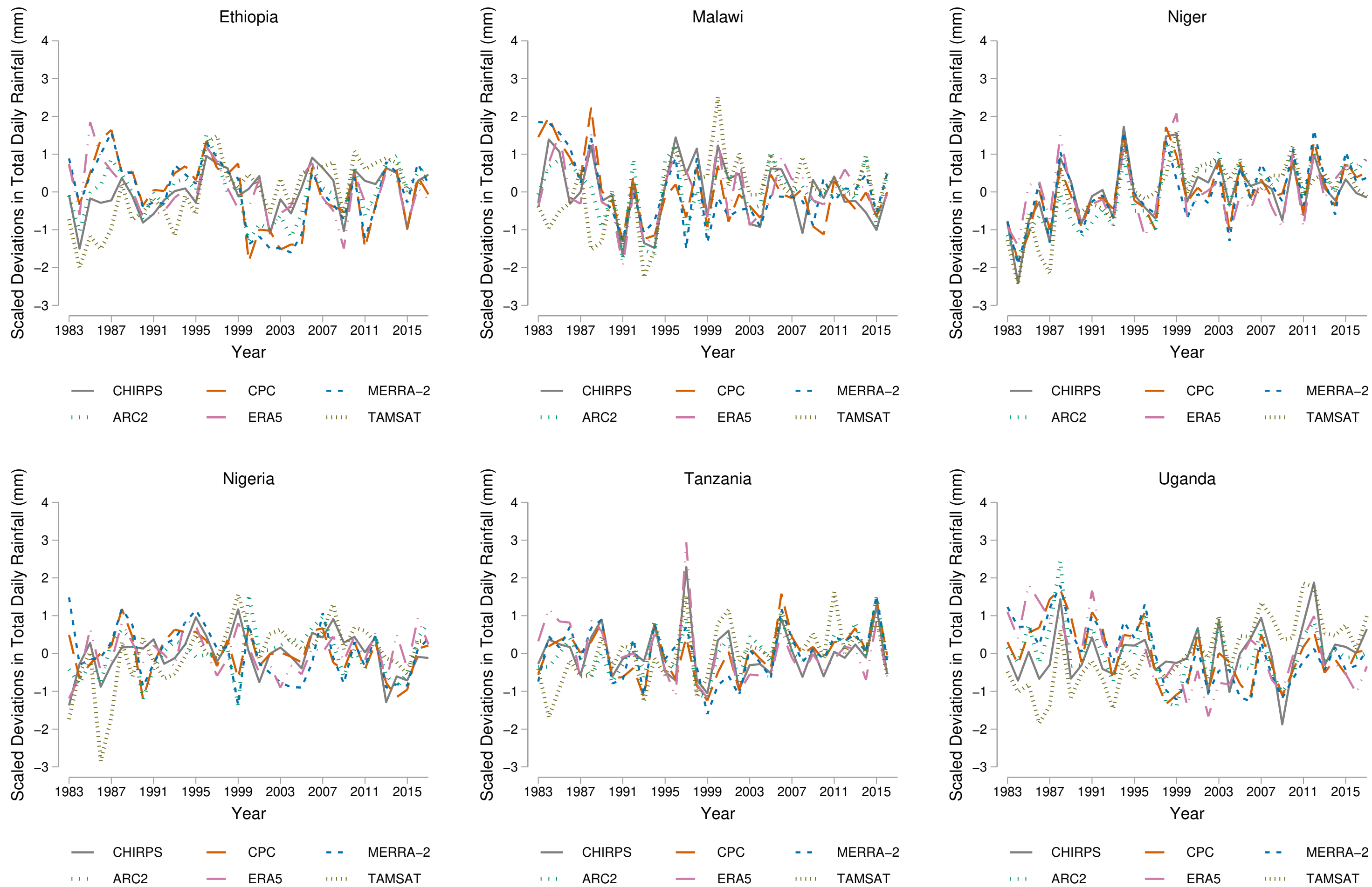
Note: The figure presents the time series of the average value of the total rainfall for each year in each country, measured by different remote sensing products. The y-axis has the total daily rainfall in mm, and the x-axis has the duration of the analysis from 1983 to current.

Figure A.5: Deviations in Total Daily Rainfall



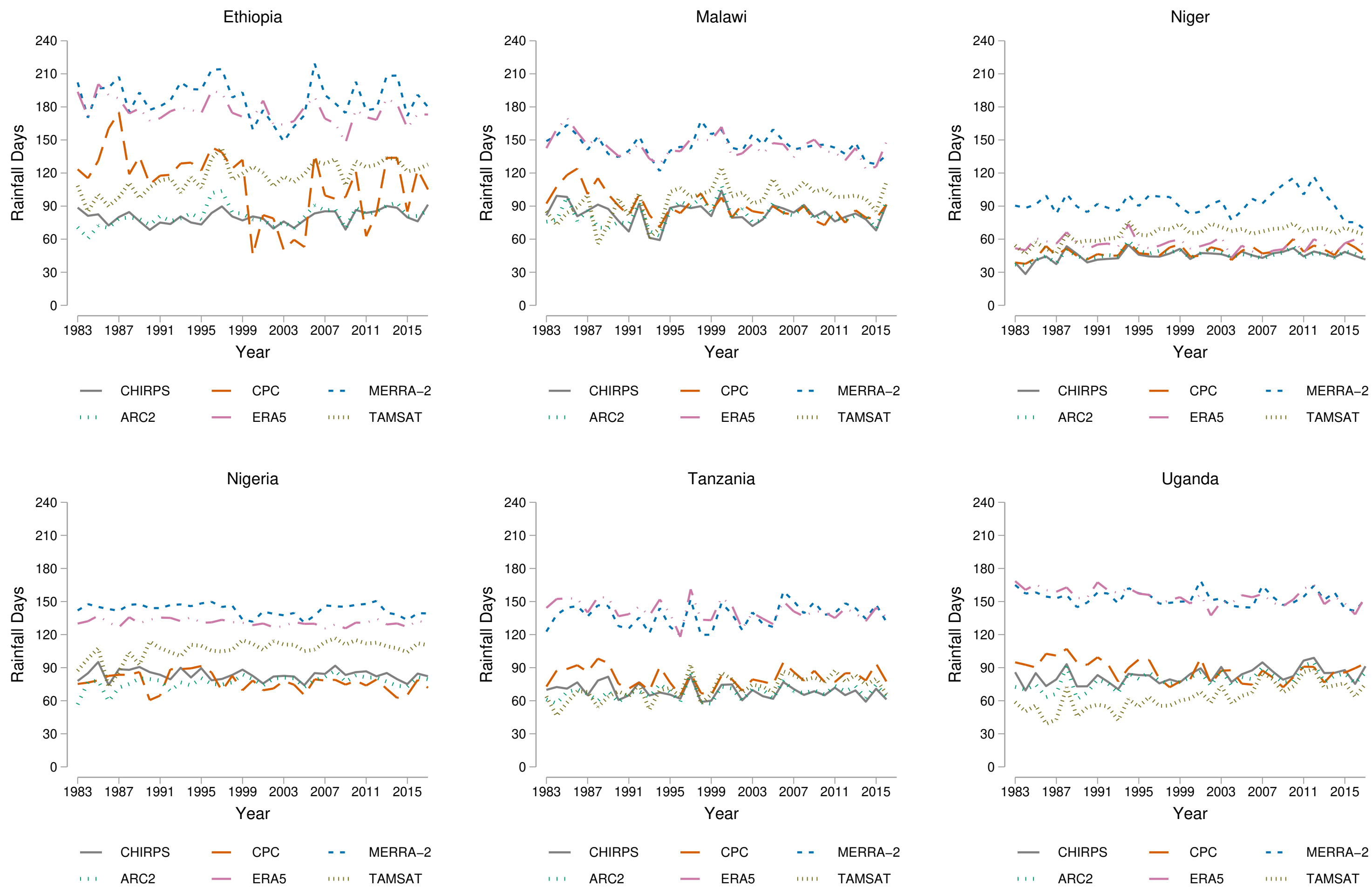
Note: The figure presents the time series of the average value of the deviations in total rainfall for each year in each country, measured by different remote sensing products. The y-axis has the deviations in total daily rainfall in mm, and the x-axis has the duration of the analysis from 1983 to current.

Figure A.6: Scaled Deviations in Total Daily Rainfall



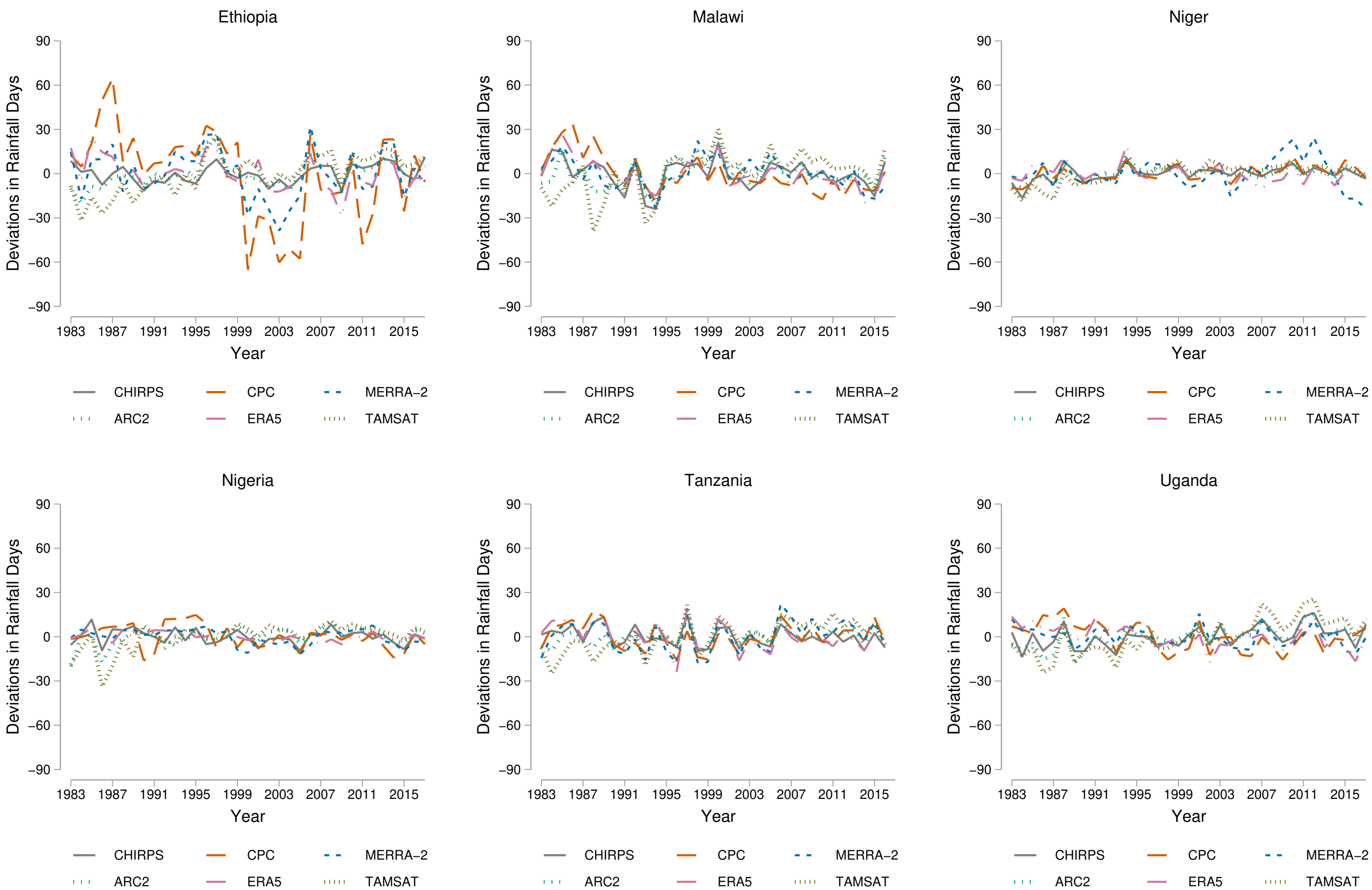
Note: The figure presents the time series of the average value of the z-score of the total rainfall for each year in each country, measured by different remote sensing products. The y-axis has the z-score of the total daily rainfall in mm, and the x-axis has the duration of the analysis from 1983 to current.

Figure A.7: Rainfall Days



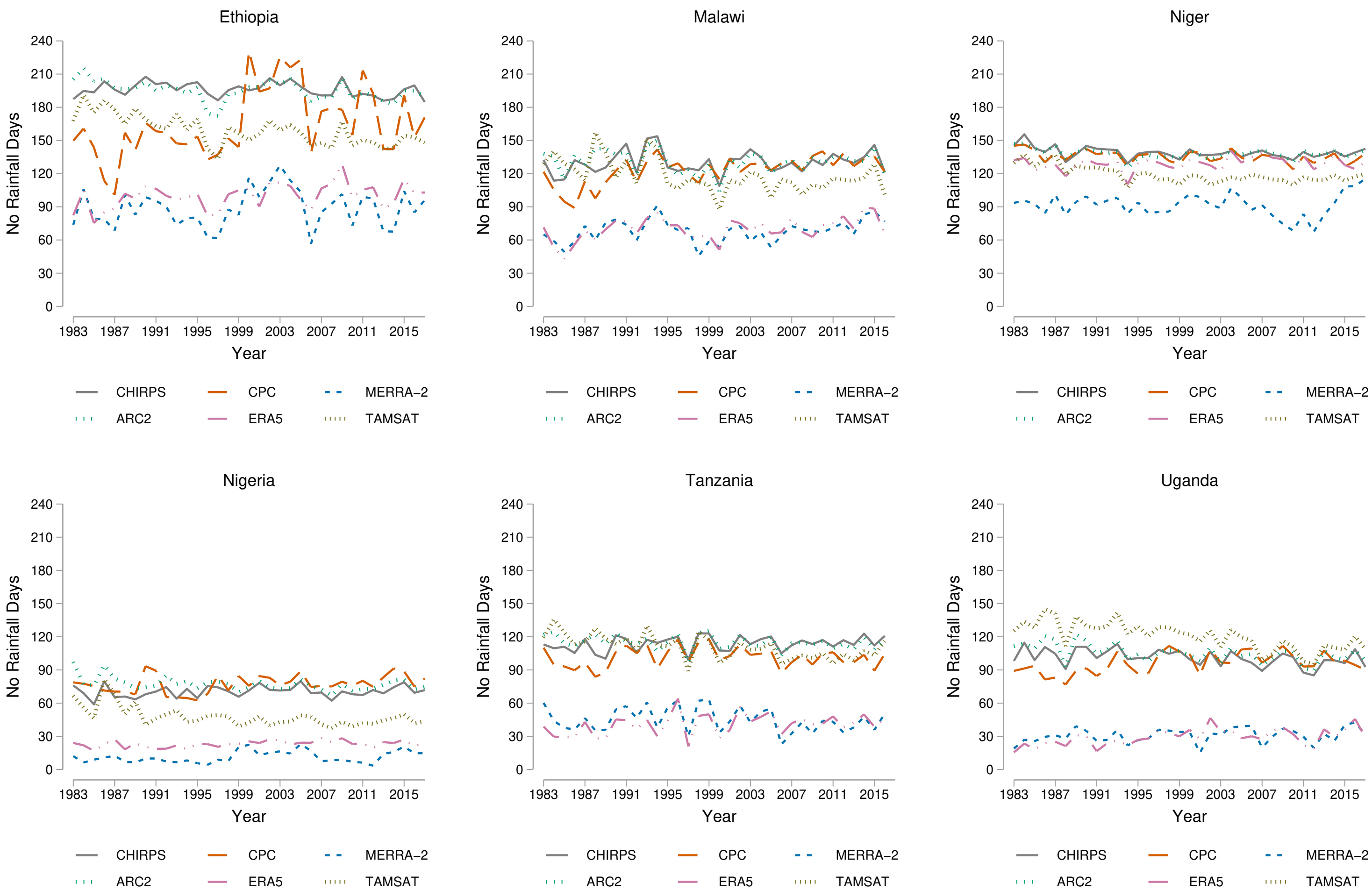
Note: The figure presents the time series of the average value of rainfall days for each year in each country, measured by different remote sensing products. The y-axis has the number of days it rained more than 1 mm in a growing season, and the x-axis has the duration of the analysis from 1983 to current.

Figure A.8: Deviations in Rainfall Days



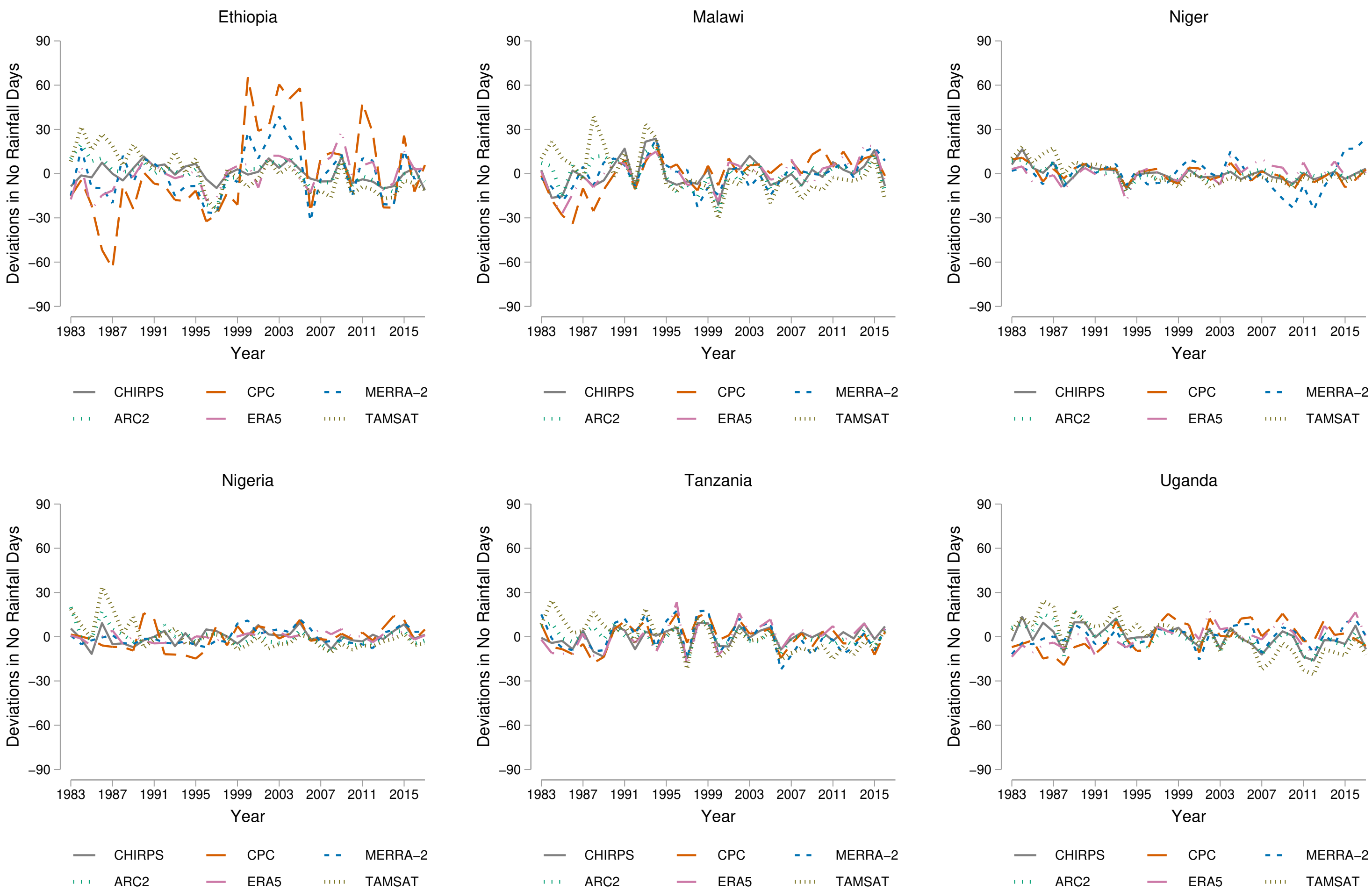
Note: The figure presents the time series of the average value of the deviations in rainfall days for each year in each country, measured by different remote sensing products. The y-axis has the deviation in rainfall days from a long-run average in a growing season, and the x-axis has the duration of the analysis from 1983 to current.

Figure A.9: No Rainfall Days



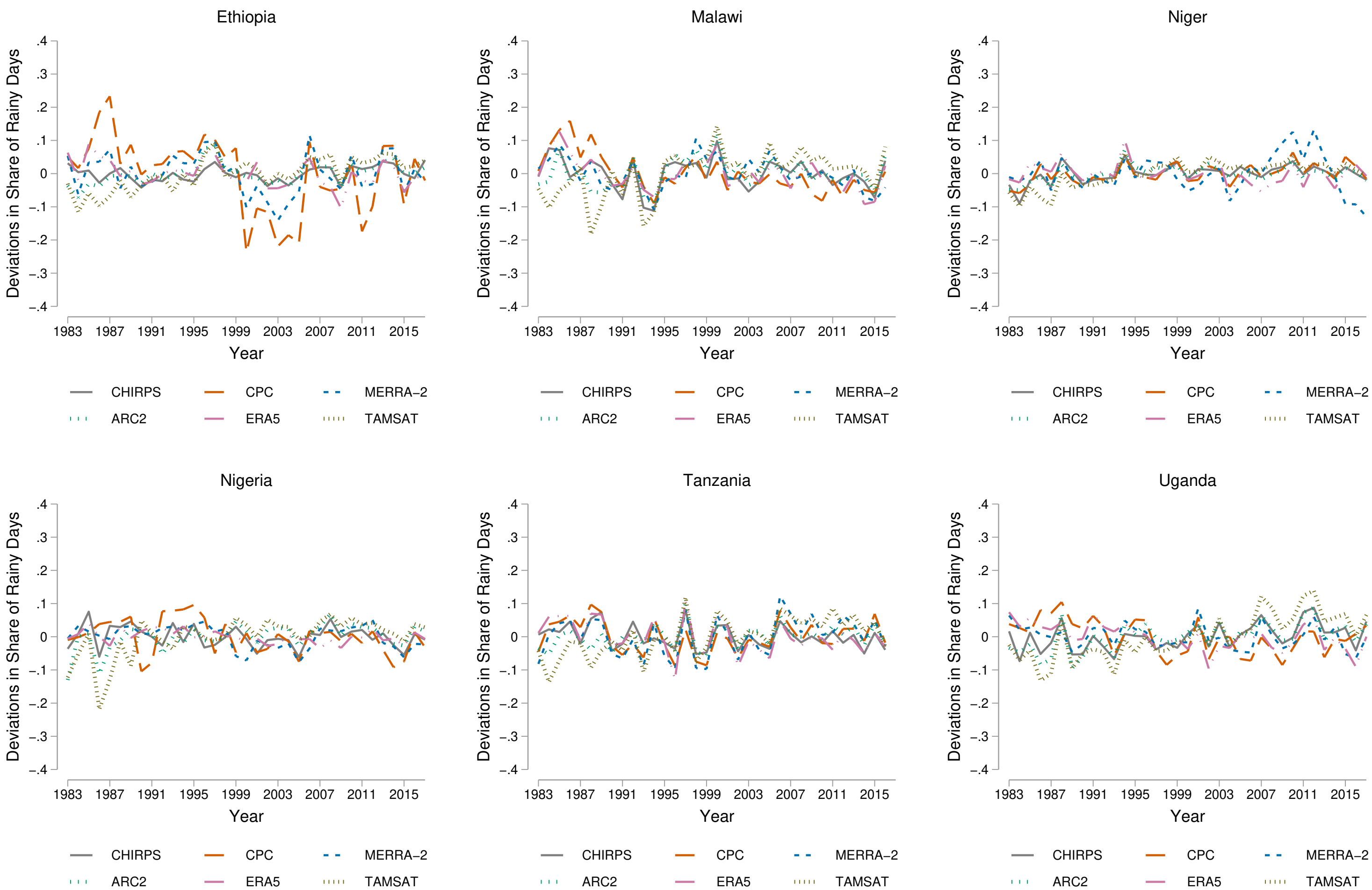
Note: The figure presents the time series of the average value of no rain days for each year in each country, measured by different remote sensing products. The y-axis has the number of days it did not rain more than 1 mm in a growing season, and the x-axis has the duration of the analysis from 1983 to current.

Figure A.10: Deviations in No Rainfall Days



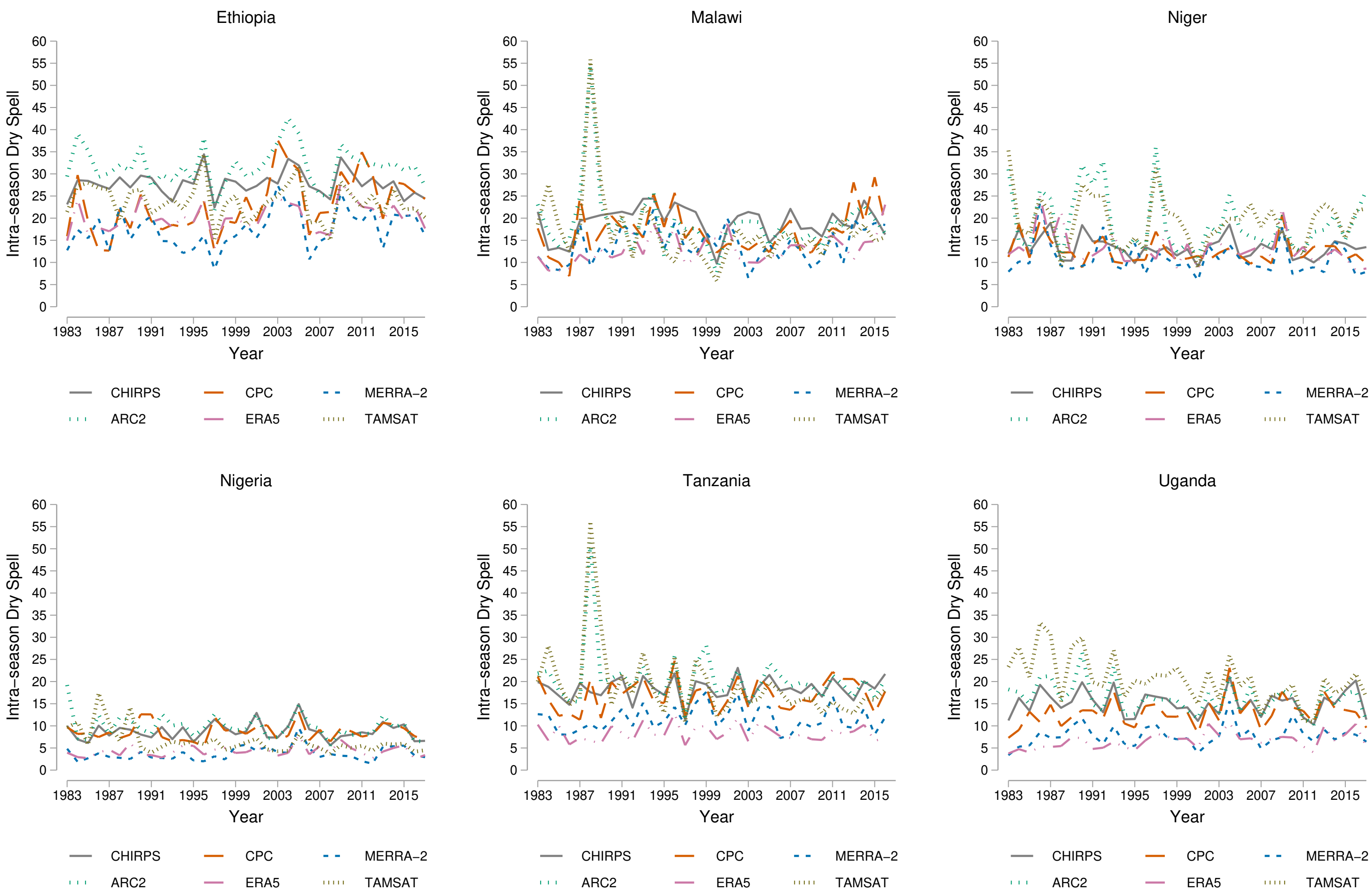
Note: The figure presents the time series of the average value of the deviations in no rainfall days for each year in each country, measured by different remote sensing products. The y-axis has the deviations in no rain days from the long-run average in a growing season, and the x-axis has the duration of the analysis from 1983 to current.

Figure A.11: Deviations in Share of Rainy Days



Note: The figure presents the time series of the average value of the deviations in the share of rainy days for each year in each country, measured by different remote sensing products. The y-axis has the deviations in the share of rainy days, and the x-axis has the duration of the analysis from 1983 to current.

Figure A.12: Intra-season Dry Spell



Note: The figure presents the time series of the average value of the intra-season dry spell for each year in each country, measured by different remote sensing products. The y-axis has the longest dry spell, and the x-axis has the duration of the analysis from 1983 to current.

REFERENCES

M. Amare, N. D. Jensen, B. Shiferaw, and J. D. Cissé. Rainfall shocks and agricultural productivity: Implication for rural household consumption. *Agric. Syst.*, 166:79–89, Oct. 2018.

I. Andrews, J. H. Stock, and L. Sun. Weak instruments in instrumental variables regression: Theory and practice. *Annu. Rev. Econom.*, (Volume 11, 2019):727–753, Aug. 2019.

J. D. Angrist and G. W. Imbens. Identification and estimation of local average treatment effects. Feb. 1995.

A. Banerjee and L. Iyer. History, institutions, and economic performance: The legacy of colonial land tenure systems in India. *Am. Econ. Rev.*, 95(4):1190–1213, Sept. 2005.

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

M. Brückner. Economic growth, size of the agricultural sector, and urbanization in Africa. *J. Urban Econ.*, 71(1):26–36, Jan. 2012.

M. Brückner and M. Gradstein. Government spending cyclicality: Evidence from transitory and persistent shocks in developing countries. *J. Dev. Econ.*, 111:107–116, Nov. 2014.

M. Brückner, W. Paczos, and E. Pappa. On the relationship between domestic saving and the current account: Evidence and theory for developing countries. *J. Money Credit Bank.*, 52(5):1071–1106, Aug. 2020.

Central Statistics Agency of Ethiopia (CSA). Rural Socioeconomic Survey 2011-2012, 2014. 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). Ethiopia Socioeconomic Survey 2013-2014, 2015. 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). Ethiopia Socioeconomic Survey, wave 3 (ESS3) 2015-2016, 2017. Public Use Dataset. Ref: ETH_2015_ESS_v02_M. Downloaded from https://microdata.worldbank.org/index.php/catalog/2783 on 6 September 2019.

J. J. Chen, V. Mueller, Y. Jia, and S. K.-H. Tseng. Validating migration responses to flooding using satellite and vital registration data. *American Economic Review*, 107(5): 441–45, 2017.

A. Deaton. Instruments, randomization, and learning about development. *J. Econ. Lit.*, 48 (2):424–455, June 2010.

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

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

F. Faradiba. The effect of rainfall on the spread of malaria in Indonesia. *International Journal of Community Medicine and Public Health*, 8(3):1146–1150, 2021.

M. Fontenla, M. Ben Goodwin, and F. Gonzalez. Pollution and the choice of where to work and live within Mexico City. *Latin American Economic Review*, 28(1):11, Nov. 2019.

C. Funk, P. Peterson, M. Landsfeld, D. Pedreros, J. Verdin, S. Shukla, G. Husak, J. Rowland, L. Harrison, A. Hoell, and J. Michaelsen. The climate hazards infrared precipitation with stations—a new environmental record for monitoring extremes. *Scientific Data*, 2(1):1–21, Dec. 2015.

A. Georgiadis, L. Benny, P. Dornan, and J. Behrman. Maternal undernutrition in adolescence and child human capital development over the life course: Evidence from an international cohort study. *Economica*, 88(352):942–968, Oct. 2021.

Ghimire Monika, Boyer Tracy A., Chung Chanjin, and Moss Justin Q. Estimation of residential water demand under uniform volumetric water pricing. *Journal of Water Resources Planning and Management*, 142(2):04015054, Feb. 2016.

W. Hanandita and G. Tampubolon. Does poverty reduce mental health? An instrumental variable analysis. *Soc. Sci. Med.*, 113:59–67, July 2014.

T. G. Hansford and B. T. Gomez. Estimating the electoral effects of voter turnout. *Am. Polit. Sci. Rev.*, 104(2):268–288, May 2010.

G. Hassan. Clean energy and household remittances in Bangladesh: Evidence from a natural experiment. Apr. 2020.

M. P. Heger, G. Zens, and M. Bangalore. Land and poverty: the role of soil fertility and vegetation quality in poverty reduction. *Environ. Dev. Econ.*, 25(4):315–333, Aug. 2020.

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

N. Hughes. Estimating irrigation farm production functions using ABARES irrigation survey data. 2011.

H. G. Jacoby and E. Skoufias. Testing theories of consumption behavior using information on aggregate shocks: Income seasonality and rainfall in rural India. *Am. J. Agric. Econ.*, 80(1):1–14, Feb. 1998.

B. F. Jones. The rise of research teams: Benefits and costs in economics. *J. Econ. Perspect.*, 35(2):191–216, May 2021.

A. Josephson and J. D. Michler. *Research Ethics in Applied Economics: A Practical Guide.* Taylor & Francis, Sept. 2023.

S. Kalemli-Özcan, A. Nikolsko-Rzhevskyy, and J. H. Kwak. Does trade cause capital to flow? evidence from historical rainfall. *J. Dev. Econ.*, 147:102537, Nov. 2020.

J. F. Kiviet. Testing the impossible: Identifying exclusion restrictions. *J. Econom.*, 218(2): 294–316, Oct. 2020.

G. Kling, U. Volz, V. Murinde, and S. Ayas. The impact of climate vulnerability on firms' cost of capital and access to finance. *World Dev.*, 137:105131, Jan. 2021.

R. Levine and D. Renelt. A sensitivity analysis of Cross-Country growth regressions. *Am. Econ. Rev.*, 82(4):942–963, 1992.

J. T. Lind. Rainy day politics. an instrumental variables approach to the effect of parties on political outcomes. *Eur. J. Polit. Econ.*, 61:101821, Jan. 2020.

A. M. Mapulanga and H. Naito. Effect of deforestation on access to clean drinking water. *Proc. Natl. Acad. Sci. U. S. A.*, 116(17):8249–8254, Apr. 2019.

J. Mellon. Rain, rain, go away: 195 potential Exclusion-Restriction violations for studies using weather as an instrumental variable. Apr. 2023.

J. D. Michler, A. Josephson, T. Kilic, and S. Murray. Empirically estimating the impact of weather on agriculture. OSF registries. July 1, 2019.

V. Munley, A. Garcia-Rodriguez, and P. Redmond. The impact of voter turnout on referendum outcomes: evidence from Ireland. *Public Choice*, 194(3):369–393, Mar. 2023.

M. S. Nasser, A. Baig, and D. Nasser. Child stunting and economic outcomes in SAARC countries: The empirical evidence. Feb. 2022.

National Bureau of Statistics (NBS). Nigeria General Household Survey (GHS), panel 2010, wave 1, 2012. 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). Nigeria General Household Survey, panel 2012-2013, wave 2, 2014. 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). Nigeria General Household Survey, panel (ghs-panel) 2015-2016, 2019. 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). Third Malawi Integrated Household Survey 2010-2011, 2012. 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). Malawi Integrated Household Panel Survey 2010-2013 (short-term panel, 204 EAs), 2015. 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). Malawi Integrated Household Panel Survey 2010-2013-2016 (long-term panel, 102 eas), 2017. Public Use Dataset. Ref: MWI_2010-2016_IHPS_v02_M. Downloaded from https://microdata.worldbank.org/index.php/catalog/2939 on 6 September 2019.

H. Ngoma, A. Finn, and M. Kabisa. Climate shocks, vulnerability, resilience and livelihoods in rural Zambia. *Climate and Development*, pages 1–12.

N. S. Novella and W. M. Thiaw. African Rainfall Climatology Version 2 for Famine Early Warning Systems. *J. Appl. Meteorol. Climatol.*, 52(3):588–606, Mar. 2013.

T. Owens, J. Hoddinott, and B. Kinsey. The impact of agricultural extension on farm production in resettlement areas of Zimbabwe. *Econ. Dev. Cult. Change*, 51(2):337–357, Jan. 2003.

P. Palacios and M. A. Pérez-Uribe. The effects of agricultural income shocks on forced migration: Evidence from Colombia. *Peace Economics, Peace Science and Public Policy*, 27(3):311–340, Sept. 2021.

J. Priem, H. Piwowar, and R. Orr. OpenAlex: A fully-open index of scholarly works, authors, venues, institutions, and concepts. May 2022.

C. Raleigh, H. J. Choi, and D. Kniveton. The devil is in the details: An investigation of the relationships between conflict, food price and climate across Africa. *Glob. Environ. Change*, 32:187–199, May 2015.

L. Rudolph. Turning out to turn down the EU: the mobilisation of occasional voters and Brexit. *Journal of European Public Policy*, 27(12):1858–1878, Dec. 2020.

X. X. Sala-i Martin. I just ran four million regressions. November 1997. URL http://www.nber.org/papers/w6252.

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

A. Sinha, Z. Shen, Y. Song, H. Ma, D. Eide, B.-J. p. Hsu, and K. Wang. An overview of microsoft academic service (MAS) and applications. In *Proceedings of the 24th International Conference on World Wide Web*, WWW '15 Companion, pages 243–246, New York, NY, USA, May 2015. Association for Computing Machinery.

J. H. Stock and M. Yogo. Testing for weak instruments in linear IV regression. Nov. 2002.

Survey and Census Division, National Institute of Statistics, Niger (NIS). National Survey on Household Living Conditions and Agriculture 2011, 2014. 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). National Survey on Household Living Conditions and Agriculture (ecvma) 2014, 2016. 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). National Panel Survey 2008-2009, wave 1, 2011. 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). National Panel Survey 2010-2011, wave 2, 2012. 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). National Panel Survey 2012-2013, wave 3, 2015. Public Use Dataset. Ref: TZA_2012_NPS-R3_v01_M. Downloaded from https://microdata.worldbank.org/index.php/catalog/2252 on 6 September 2019.

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

T.-A. Trinh, S. Feeny, and A. Posso. Rainfall shocks and child health: the role of parental mental health. *Climate and Development*, 13(1):34–48, Jan. 2021.

T.-A. Trinh, P. Srivastava, and S. Brown. Household expenditure and child health in Vietnam: analysis of longitudinal data. *Journal of Demographic Economics*, 88(3):351–377, Sept. 2022.

Uganda Bureau of Statistics (UBOS). Uganda National Panel Survey (UNPS) 2010-2011, 2014a. 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). Uganda National Panel Survey (UNPS) 2010-2011, 2014b. 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). National Panel Survey (UNPS) 2005-2009, 2019. Public Use Dataset. Ref: UGA_2005-2009_UNPS_v01_M. Downloaded from https://microdata.worldbank.org/index.php/catalog/1001 on 6 September 2019.

S. Veljanoska. Do remittances promote fertilizer use? the case of Ugandan farmers. *Am. J. Agric. Econ.*, 104(1):273–293, Jan. 2022.