

The role of statistical distributions in vulnerability to poverty analysis

Armine Poghosyan

Thesis submitted to the Faculty of the
Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of

Master of Arts
in
Data Analysis and Applied Statistics

Jennifer Van Mullekom, Chair

Elinor Benami

Klaus Moeltner

April 11, 2024

Blacksburg, Virginia

Keywords: satellite remote sensing, gridded weather, consumption.

Copyright 2024, Armine Poghosyan

The role of statistical distributions in vulnerability to poverty analysis

Armine Poghosyan

ABSTRACT

In regions characterized by semi-arid climates where households' welfare primarily relies on rainfed agricultural activities, extreme weather events such as droughts can present existential challenges to their livelihoods. To mitigate these risks, numerous social protection programs have been established to assist vulnerable households affected by weather events. Despite efforts to monitor environmental changes through remotely sensed technology, estimating the impact of weather variability on livelihoods remains challenging. This is compounded by the need to select appropriate statistical distribution for weather anomaly measures and household characteristics. We address these challenges by analyzing household consumption data from the Living Standards Measurement Study survey in Niger and systematically evaluating how each input factor affects vulnerability estimates. Our findings show that the choice of statistical distribution can significantly alter outcomes. For instance, using alternative statistical distribution for vegetation index readings could lead to differences of up to 0.7%, which means around 150,000 more households might be misclassified as not vulnerable. Similarly, variations in household characteristics could result in differences of up to 10 percentage points, equivalent to approximately 2 million households. Understanding these sensitivities helps policymakers refine targeting and intervention strategies effectively. By tailoring assistance programs more precisely to the needs of vulnerable households, policymakers ensure that resources are directed where they can make the most impact in lessening the adverse effects of extreme weather events. This enhances the resilience of communities in semi-arid regions.

The role of statistical distributions in vulnerability to poverty analysis

Armine Poghosyan

GENERAL AUDIENCE ABSTRACT

In drought-prone regions where many families rely on rainfed farming, extreme weather can devastate livelihoods. Governments have created aid programs to assist the most vulnerable households during these climate crises, but identifying who needs help is extremely challenging. Part of this difficulty lies in selecting the right statistical methods for analyzing weather data and household information. In this paper, we focus on Niger, a country that experiences frequent droughts and where over 80% of the population depends on rainfed agriculture. By evaluating household consumption data, we aim to assist in identifying the households who has high probability of becoming poor as a result of unfavorable weather events and thus needs support from social protection programs. In our analysis, we systematically evaluate how each input factor (including household characteristics and statistical distributions) affects households likelihood of becoming poor in the event of weather crises. We find that compared to alternative statistical distributions, using a conventional normal distribution could lead to misclassifying around 150,000 households as non-vulnerable, leaving them without vital assistance. Similarly, using different sets of household characteristics can result in up to 10 percentage points which equivalents to 2 million households that would miss out on much-needed support. Understanding these sensitivities is crucial for policymakers in refining how aid programs identify the vulnerable populations and include them into the protection programs. The improved targeting approach will enhance the resilience of communities in semi-arid regions facing increasing weather variability.

Acknowledgments

I would like to express my sincere gratitude to my thesis advisor, Dr. Jennifer Van Mullekom, for her invaluable guidance, support, and encouragement throughout this journey. I am also grateful to my committee members, Dr. Elinor Benami and Dr. Klaus Moeltner, for their insightful feedback, constructive criticism, and valuable suggestions, which have greatly enriched this work. Additionally, I am thankful to my family and friends for their unconditional love, encouragement, and understanding. Their unwavering support has been a constant source of strength and motivation.

Contents

List of Figures	vii
List of Tables	ix
1 Introduction	1
2 Context	5
3 Data and Summary Statistics.	6
3.1 Household-Level Consumption	6
3.2 Weather anomaly proxies	7
3.2.1 Precipitation	8
3.2.2 Normalized Difference Vegetation Index (NDVI)	9
3.2.3 Standardized measures (z-score)	9
3.3 Combining spatial and survey data	10
4 Methods	10
4.1 Estimating vulnerability to poverty	11
4.1.1 Estimating mean consumption	11
4.1.2 Estimating variance from cross-sectional data via simulation	12
4.1.3 Calculating probability of falling into poverty	13
4.1.4 Determining household vulnerability as a function of its probability of falling into poverty	13
4.2 Robustness of vulnerability estimates to changing input assumptions	13

4.2.1	Examining the impact of historical weather data distribution on vulnerability	14
4.2.2	Assessing additional household characteristics as potential mediators	16
5	Results and Discussion	17
5.1	Vulnerability Measure	17
5.2	Sensitivity of vulnerability estimates to changing input assumptions	19
5.2.1	Results for choice of historical distribution	19
5.2.2	Results for additional household characteristics included in the analysis . . .	22
6	Conclusion	22
	Bibliography	24

List of Figures

1	Livelihood Systems in the Sahel	28
2	Average annual per capita consumption (in dollars) per each level 2 administrative unit (department) expressed in 2011 prices. Across all years, the north-western and south-western (around the capital city) portions of Niger tend to have higher average p.c. consumption.	29
3	Average precipitation and NDVI from 1982-2019, mean aggregated over the department.	30
4	Distribution of annual cumulative precipitation (mm) in Niger for the period of 2000 - 2019. Overall, there is not much variation in cumulative precipitation across time. However, within a country across all specified years, the northern regions of Niger receive lower precipitation (due to being close to the Sahara desert) compared to the southwestern regions (due to being close to the savannas).	31
5	Distribution of peak NDVI in Niger for the period of 2000 - 2019. Overall, the northern regions of Niger record lower NDVI measures compared to the ones in the southern regions. This pattern is expected considering that the northern part of the country neighbors a desert, while the southern regions neighbors the savannas.	32
6	Empirical densities of all selected weather condition measures from the combined dataset. The histograms illustrate that the anomaly measures deviate from normal distribution.	33
7	Cullen and Frey graph for all selected weather anomaly measures. It shows that the logistic distribution fits well for NDVI and its standardized measures. However, for rainfall, the potential candidates for distribution are Gamma and Log-normal. A logistic distribution might be more suitable for the standardized measure of rainfall.	34
8	Integrated visual analysis of cumulative rainfall : distribution comparison and probability plots. This figure shows that Exponential, Weibull, and Gamma distributions provide a comparatively better fit for the cumulative rainfall data.	35

9	Integrated visual analysis of standardized measure of rainfall (z-score) : distribution comparison and probability plots. This figure shows that Logistic distribution provides a comparatively better fit for the data.	36
10	Integrated visual analysis of vegetation index : distribution comparison and probability plots. This figure shows that except for the Exponential distributions, almost all other selected ones provide a good fit for the data.	37
11	Integrated visual analysis of standardized measure of vegetation index (z-score) : distribution comparison and probability plots. This figure shows that Logistic and Normal distributions provide a similar fit for the data.	38
12	Summary of vulnerability results from ~20,500 estimated models. For the majority (73%) of the estimated models, the average vulnerability estimate is between 31 - 32%.	39
13	Results of vulnerability analysis. We present the results from randomly selected 50 models out of the pool of 20,500 estimated models.	40
14	Results of vulnerability analysis for rainfall data.	41
15	Results of vulnerability analysis for the standardized measure of rainfall.	42
16	Results of vulnerability analysis for NDVI.	43
17	Results of vulnerability analysis for the standardized measures of NDVI.	44
18	Average vulnerability estimate across space. The most vulnerable non-poor rural households reside in the south and south-western part of Niger.	45

List of Tables

1	Descriptive Statistics of LSMS Data (in 2011 US\$, PPP)	46
2	Descriptive Statistics for Agro-environmental Data	47
3	Definition and Summary Statistics of Household Characteristics	48
4	Summary of the findings from the goodness of fit assessment.	49

1 Introduction

The Sahel, a vast semi-arid region transition area bordering the Sahara desert to the north and the tropical Savannas to the south¹, is known for its fragile ecosystems vulnerable to climate change impacts ([Food & Agriculture Organization, 2019](#)). With some of the highest population growth rates globally, ranging from 2.5% to 4% ([Population Reference Bureau, 2015](#)) annually, Sahelian countries are projected to double their population growth rate between 2020 and 2050 ([United Nations, 2019](#)). Escalating population growth directly translates to increased food demand within the Sahel. Such a surge in population intensifies pressure on already strained natural resources and agricultural systems, exacerbating food security issues in the region.

Apart from escalated population growth rates, the Sahel experiences frequent and severe adverse weather events, such as droughts, which can devastate the region's crops, livestock, and livelihoods ([Hallegatte et al., 2016](#)). The Sahel is typically considered arid and hot, with average annual temperatures ranging from 74 - 95°F ([Lu, 2019](#)), and strong seasonal variations in rainfall across time and space. In particular, the southernmost regions (Figure 1) of the Sahel tend to receive the most precipitation (average 300 - 500 mm annually), with most falling between June through September, whereas the northern regions receive much less (average 100 - 300 mm annually). Since the mid-1960s, the annual rainfall in the region has declined by over 30%, and correspondingly with more frequent drought events occurring in the region ([Held, 2000](#); [Mortimore, 2010](#)). Future trends suggest that extreme weather events are expected to increase in intensity and duration, affecting not only food production but also market dynamics, potentially resulting in food shortages, price volatility, and household livelihoods ([Mirza, 2003](#)). Smallholder farmers, who form the backbone of food production in the Sahel, face mounting difficulties in meeting the rising food demand amidst unpredictable weather patterns and climatic variability. These challenges are even more pronounced for farmers reliant on rain-fed agriculture, amplifying their vulnerability to fluctuations in rainfall and exacerbating food insecurity in the region ([Food & Agriculture Organization, 2019](#)). Hence, understanding the impact of weather events on household livelihoods is essential for formulating sustainable and effective programs aimed at enhancing household resilience to such adverse weather events.

Prior research has extensively investigated the relationship between extreme weather conditions, such as droughts and floods, and the welfare of rural households in developing countries ([Gao and Mills, 2018](#);

¹The definition of which countries are counted as within the Sahel varies, but it generally includes at least the arid and semi-arid areas spanning Niger, Burkina Faso, Mali, Chad, Senegal, and Mauritania ([World Bank, 2020b](#)). Another definition is circumscribed by coordinates, e.g., 12°-20°N, 17°W-40°E ([Heinrigs, 2010](#)).

Kurosaki, 2015; Lohmann and Lechtenfeld, 2015; Mera, 2018; Nguyen and Nguyen, 2020; Reynaud et al., 2018; Twongyirwe et al., 2019). These rural households were found to be particularly vulnerable to adverse weather events due to their initial low welfare levels (Nguyen and Nguyen, 2020). Research across various countries, including Mexico, Tanzania, Ethiopia, Zimbabwe, Vietnam, and others, consistently demonstrated that extreme weather variability leads to reduced household welfare, measured as income or per capita consumption (Christiaensen and Sarris, 2007; Hirvonen, 2016; Skoufias and Vinha, 2012).

For instance, in Mexico, Skoufias and Vinha (2012) reveal that deviations from expected temperature and rainfall patterns had a negative impact on household consumption. However, the magnitude of the impact varied depending on the nature of the weather shock and the specific agroecological region under investigation. Christiaensen and Sarris (2007) show that drought shocks in Tanzania lead to a decrease in household per capita consumption expenditure. Similarly, in the same context of Tanzania, Hirvonen (2016) finds that increasing temperature reduces household welfare. Moreover, literature also shows that rainfall failures and prolonged droughts impact income, consumption, and health outcomes in Ethiopia (Dercon et al., 2005; Porter, 2012) and Zimbabwe (Hoddinott, 2006). In Vietnam, household welfare is mainly affected by floods, storms (Arouri et al., 2015), and hurricanes (Skoufias et al., 2019; Thomas et al., 2010).

In essence, prior studies emphasize a concerning reality: adverse weather conditions significantly and negatively affect household welfare, especially in developing countries, where a large portion of the population relies on agriculture. This underscores the urgency to assess and implement focused interventions for improved household resilience.

In response to the negative impact of adverse weather events on household welfare, a number of programs have emerged in efforts to break the link between extreme environmental conditions and household welfare crises. These programs aim to assist households in preventing the effects of adverse weather from materializing by providing aid quickly to affected areas. Additionally, some of these programs have relied on measures of precipitation deficit or other proxies of anomalous environmental conditions, such as severe drought indicators and/or abnormal vegetation index readings, to direct their efforts. One recent example of such a program is a national safety net program implemented by the Government of Niger. This program relies on satellite-based estimates of precipitation and evapotranspiration to determine a household's eligibility for assistance, and as of last year has disbursed cash payments to more than 15,000 households affected by droughts (Brunelin et al., 2022; Mahamadou and Murthi, 2022).

The increasing availability of data about environmental conditions *and* recent advances in data availability at the household level provide important new opportunities for understanding the links between weather anomalies and household welfare. With the emergence of repeated household consumption level data from World Bank-supported initiatives such as the Living Standards and Measurement Surveys (LSMS), for example, combined with publicly available measures of environmental stress at ever-finer spatial resolutions, we now possess the ability to gain greater insight into the relationship between weather and vulnerability across time and space.

Nonetheless, estimating vulnerability to poverty measures remains complex. Typically, vulnerability assessments rely on measuring households' welfare through their consumption level, often using the poverty line as a benchmark. While many ways to quantify vulnerability exist, nearly all rely on calculating the expected value and variance of per capita consumption (a proxy to household welfare). The literature differentiates the process of estimating the variance of per capita consumption depending on the available data type. Estimating the variance is more straightforward when utilizing panel data, which contains multiple observations of the same household's consumption over time. However, in many developing countries, such as countries in the Sahelian region, nationally representative data is cross-sectional ([Baquie and Fuje, 2020](#)). In such cases, a widely used approach for estimating the variance is through simulated consumption data (see for instance [Skoufias et al. \(2019\)](#), [Baquie and Fuje \(2020\)](#)).

When estimating the variance of consumption data, and hence vulnerability measure, literature (see for example, [Baquie and Fuje \(2020\)](#); [Skoufias et al. \(2019\)](#)) typically imposes a specific distributional assumption on selected weather condition measures. Most commonly, the assumption is that agro-environmental/weather conditions follow a normal distribution, which, however, may not reflect reality. For instance, by assuming a normal distribution for rainfall, we suggest that it is evenly distributed year-round. However, in reality, the Sahelian region receives most of its rainfall during the growing season, spanning from June to September, with little to no precipitation during other months. This mismatch between assumed and actual distribution can result in inaccurate vulnerability estimates. Moreover, deciding which household-level characteristics to incorporate into the estimation framework is challenging, as not all characteristics may carry equal importance in assessing vulnerability. Adding irrelevant or less significant factors could complicate the identification of the primary drivers of vulnerability estimates.

In this work, we thus seek to address two main questions:

1. How do selected statistical distributions for weather anomaly measures influence the vulnerability estimate?
2. How does the choice of household characteristics influence the relationship between weather and vulnerability?

Our research draws on the literature on weather-induced vulnerability to poverty to estimate the relationship between weather and household welfare proxies. We then extend this work by systematically evaluating how factors such as distributional assumptions of input data, i.e., proxies for weather conditions affect the sensitivity or stability of vulnerability estimates. In particular, we estimate the relationship of environmental conditions – proxied by a time series of spatially explicit drought indicators, including precipitation deficit and vegetation index – on household total consumption expenditure extracted from the World Bank-supported LSMS survey across all available years. We then build upon these estimates to evaluate which households appear most vulnerable to weather variations, building on vulnerability assessment methods used by [Baquie and Fuje \(2020\)](#); [Chaudhuri \(2000\)](#); [Hill and Porter \(2017\)](#); [Skoufias et al. \(2019\)](#), who provide vulnerability estimates for the respective study areas based on a series of relevant inputs but often do not systematically examine the sensitivity of their results as a function of reasonable alternative decisions on input variables. Thus, one of our key contributions to this literature revolves around assessing the sensitivity of our vulnerability analysis results based on varying reasonable input decisions.

We find that depending on the weather condition measure under consideration, the vulnerability estimates can be sensitive to the choice of distributional assumptions. Specifically, the vegetation index proves to be more sensitive to the correct statistical distributional assumption, resulting in up to a 0.74% difference in vulnerability estimates. While this percentage change may appear insignificant at first glance, it signifies a potential misclassification of approximately 150,000 additional households if we adhere to the Normal distribution assumption.

Additionally, including household characteristics beyond size leads to a significant 10 percentage point change in the estimate. Specifically, we find that household size, household head age and sex, household living conditions (i.e., variables describing the main construction materials for floor, walls, and roof), and the status of the ownership of a car are the key drivers of household vulnerability to poverty estimates. Overall, our findings underscore the need for careful consideration of these factors in developing social protection programs.

The remainder of this paper is organized as follows: Section 2 provides relevant context on the region of interest. Section 3 characterizes the socioeconomic as well as geospatial data sources used in the analysis. Section 4 outlines the methods used to examine vulnerability, Section 5 summarizes the results, and Section 6 concludes with a view towards what types of indicators seem to connect most closely to consumption, and what these analyses portend for the next generation of drought relief programs.

2 Context

This study focuses on Niger, a country emblematic of the challenges faced by households in the Sahel region. With a staggering 43% poverty rate ([Concern Worldwide U.S., 2022](#)) and a population of 25.4 million ([Central Intelligence Agency, 2024](#)), Niger stands among the world's poorest countries. The country's climate, marked by consistently high temperatures ranging from 72 to 97°F annually ([World Bank Group Climate Change Knowledge Portal, 2021](#)), is characterized by a prolonged, intense dry season from October to May, succeeded by a brief, erratic rainy season spanning June to September ([World Bank Group Climate Change Knowledge Portal, 2021](#)). Notably, annual rainfall displays significant variation, with the north experiencing lower levels (100–200 mm) compared to the more abundant rainfall in the south (500–600 mm) ([World Bank Group Climate Change Knowledge Portal, 2021](#)).

This climatic diversity significantly shapes the economic activities of Niger's households. In the north, pastoralist livelihoods predominate, while rainfed agriculture takes precedence in the south. Rainfed agriculture holds particular economic significance, contributing up to 40% to the country's Gross Domestic Product and serving as the livelihood for nearly 80% of the population dependent on this form of agriculture ([Brunelin et al., 2022](#)).

However, the fragility of agriculture and pastoralism in Niger is aggravated by the recurrent threat of droughts. Frequent droughts, a longstanding challenge for the region, have historical ties to triggering large-scale famines and excess mortality in Sahelian countries, as observed in the periods 1968-1974 and 1982-1983 ([Mortimore, 2010](#); [Sen, 1982](#)). More recently, Niger experienced severe droughts leading to food security crises in 2005, 2008, 2010, and 2012 ([Brunelin et al., 2022](#)). These weather events not only exacerbate poverty in the country but also amplify the vulnerability of households to fluctuations in weather patterns. This highlights the significance of investigating the connection between household consumption and weather events to formulate efficient strategies to alleviate their effects on vulnerable households.

3 Data and Summary Statistics

To estimate household-level vulnerability to poverty associated with weather conditions, we combine data on per capita consumption with indicators of agricultural conditions, including precipitation and vegetation index, often used as proxies for vegetative stress. This section describes these datasets and how we merge environmental indicators with household-level consumption data.

3.1 Household-Level Consumption

To proxy household welfare, we follow the previous literature ([Baquie and Fuje, 2020](#); [Hill and Porter, 2017](#); [Skoufias et al., 2019](#)) and use per capita household level consumption, here calculated as the sum of total food and non-food expenditures in a given household throughout the survey year divided by the number of household members. The household-level expenditure data come from three waves of the Nigerien/World Bank LSMS. National statistical agencies are responsible for implementing these LSMS surveys. Although each implementing country may vary in which questions they choose to include, the basic structure of LSMS was initially designed by the World Bank to help partner countries capture nationally representative statistics on household living conditions, welfare, income, consumption, and a wide variety of household socio-economic characteristics ([World Bank, 2020a](#)). The resulting survey data include geographic flags that allow us to observe approximately from where the observations are drawn. Specifically, we are provided the latitudes and longitudes of enumeration areas, reflecting a centroid from multiple interviewed households. We use this centroid to determine which administrative units to assign to each household via a spatial overlap (join) between the centroid's coordinates and the administrative boundaries for each department (administrative unit 2). To further evaluate the localized impact of weather anomalies on per capita household consumption, we created circular buffers with a 20 km radius around the geo-referenced locations or enumeration areas using the same centroids obtained from multiple interviewed households.

Thus, although the survey data reflect household responses resolved at the household level, location data are approximate, and some errors may arise from the jitter associated with using the enumeration area values and the spatial join to an administrative unit. Although we are aware that LSMS surveys are designed to capture nationally (and not sub-nationally) representative details, they remain the best-known repeated source of information on socio-economic conditions that include some information on subnational data available to the public. With that in mind, this study employs data from Niger for all available LSMS years, i.e., 2011,

2014, and 2018, for a total of 12,832 observations across the three years.

Across all years, the north-western and south-western areas of Niger tend to exhibit higher annual per capita consumption than other regions, with the areas surrounding the capital city of Niamey exhibiting relatively high consumption in each year (see Figure 2). We can explain this trend by several key factors. Firstly, the north, being involved in uranium extraction, likely experiences higher levels of economic welfare due to the revenues generated from this industry (Sisa, 2014; World Nuclear Association, 2023). This could lead to overall higher purchasing power and higher consumption levels. Additionally, Niamey, as the capital city, likely serves as a hub for commerce and trade, further boosting consumption in the surrounding areas. Conversely, the relatively lower consumption levels in the southern regions may be attributed to their heavy reliance on agriculture, which is significantly impacted by frequent climatic shocks such as droughts (Central Intelligence Agency, 2024; Sisa, 2014). This reliance on agriculture makes households living in the southern regions more vulnerable to fluctuations in income, thereby limiting their purchasing power and resulting in comparatively lower consumption levels.

Table 3 provides a snapshot of the surveyed households. On average, a typical household comprises six members, headed by a 45-year-old male, with 87% having no formal education. Most households (68%) practice monogamous marriages and live in dwellings made of earth-based materials (roof: 31%, wall: 46%, floor: 95%). Basic amenities are lacking, with 97% lacking electricity, 98% lacking sanitation facilities, and most (around 99.7%) not owning cars. These households, on average, possess about nine livestock.

Our analysis focuses on rural areas where households are engaged in rainfed agricultural activities. To ensure the reliability of the regression analysis presented in Section 4, we refine our dataset by retaining only those administrative units where we have at least 30 data points across all the survey years. This results in a total of 8,425 observations.

3.2 Weather anomaly proxies

Although a variety of indicators can reflect weather conditions, we draw upon two primary indicators: precipitation and a proxy of vegetative stress. In this study, we aggregate all weather condition proxy variables to match the spatial units at which consumption data is available. Specifically, we calculate the average of all values across each geographic unit, i.e., a 20km buffer surrounding the stated enumeration area, over the year. We construct these weather condition proxies using the growing season preceding the

household survey month in Niger, determined based on crop-specific harvest calendars from the European Union’s Joint Research Council. We collect all agro-environmental variables for the period spanning from 1982 to 2019. The remainder of this section provides a comprehensive description of these measures and their respective data sources.

3.2.1 Precipitation

Annual precipitation data comes from the Climate Hazards Group Infrared Precipitation With Station (CHIRPS) dataset. CHIRPS was released in 1981 as a means to provide reliable, low-latency seasonal drought monitoring data for the Famine Early Warning Systems Network (FEWS NET) (Funk et al., 2015), and now it is a commonly used measure of precipitation across the globe. Precipitation in CHIRPS is recorded daily at a 5 x 5 km spatial resolution. Using the daily CHIRPS rainfall data, we construct cumulative rainfall of the prior to household survey growing season for one month period. In addition to considering the cumulative precipitation (measured in mm), we compiled data on seasonal precipitation. This step was undertaken due to the Sahelian countries’ susceptibility to extreme aridity and their significant dependence on rainfall, specifically during the cultivation season, which spans from June to September (Climates to Travel, 2023; Funk et al., 2012; Weather Spark, 2023). We constructed seasonal precipitation values by aggregating the daily precipitation data into monthly values for the June-September period. Panel A of Figure 3 and Table 2 presents the national annual mean precipitation for the years of interest (2011, 2014, 2018) relative to the values from the 29-year time period preceding our analysis. Notably, the year 2018 stands out with above-average rainfall, while 2011 experienced the lowest rainfall of the three years in our study. The low precipitation in 2011 resulted in a drought in the area, which later translated into food insecurity crises in 2012 (Brunelin et al., 2022).

Examining the distributions of cumulative annual precipitation within Niger for the period 2000-2019, presented in Panel A of Figures 4, shows that the northern regions of Niger receive lower annual precipitation compared to the southwestern parts. This finding is in line with expectations, as the northern areas are covered by the Sahara desert, while the southern parts are closer to the African savannas. Despite these geographical differences, however, we did not observe much variation in the annual cumulative precipitation over time. The histograms in Panel B of Figure 4 further support our findings, indicating a consistent distribution of annual precipitation across the years 2000-2019 ².

²Insufficient variation in the weather proxies can weaken our ability to estimate the relationship between the consumption

3.2.2 Normalized Difference Vegetation Index (NDVI)

We use the Normalized Difference Vegetation Index (NDVI) derived from measurements taken by the Advanced Very High-Resolution Radiometer (AVHRR) space-borne sensor at 1 km spatial resolutions. The NDVI measure captures the difference between visible and near-infrared reflectance of a given area calculated as:

$$NDVI = \frac{R_{NIR} - R_{RED}}{R_{NIR} + R_{RED}}$$

where R_{NIR} represents the light reflected in the near-infrared spectrum, and R_{RED} indicates the light reflected in the red range of the spectrum.

The formula above measures NDVI as a dimensionless index, with potential values ranging from -1 to 1. The closer NDVI is to 1, the more the area is associated with dense vegetation and/or crops at the peak of their growth. Usually, negative values of NDVI represent water bodies. Values between -0.1 and 0.1 are related to barren areas, rock, sand, or snow. NDVI values in the 0.2 to 0.4 range represent shrub and grassland, while higher (around 0.6 - 0.8) values indicate forest areas (Weier and Herring, 2000). Similar to precipitation, we construct the environmental condition measures for NDVI using the peak NDVI readings from the preceding growing season.

Consistent with the findings regarding rainfall data, the peak NDVI measure exhibits higher values in the southern regions, where agricultural activities are predominantly concentrated (refer to Figure 5). Additionally, when comparing the NDVI values across the years of interest—2011, 2014, and 2018—we observe that, on average, 2018 records a higher NDVI value (see Figure 3). This suggests that 2018 experienced healthier vegetation growth than the other years, possibly due to favorable weather conditions and abundant rainfall.

3.2.3 Standardized measures (z-score)

Along with cumulative annual and seasonal rainfall and peak NDVI, we also use standardized measures for each agro-environmental variable. These standardized measures represent how many standard deviations each annual value deviates from the historical mean, which is defined using the 1982-2010 period. We

and weather conditions accurately. This limitation may lead to inflated standard errors, reduced predictive power, and potential underestimation of the true effects of the weather variations on the per capita household consumption.

calculate standardized anomalies using the following formula:

$$W_{it} = \frac{w_{it} - \bar{w}_i}{s_i} \quad (1)$$

where w_{it} is the value of the agro-environmental indicator for a given area i and a given year t ; \bar{w}_i is the sample mean across the baseline period for area i ; s_i is the standard deviation of the sample across the baseline period for area i .

3.3 Combining spatial and survey data

To merge the weather data with the LSMS survey data, we filter the weather measures for the years 2011, 2014, and 2018, matching the survey data years. This process yields three observations for each location within a 20km buffer around the enumeration area, one for each of the specified years of interest. Finally, we merge the survey data with the weather measures based on the year and unique location ID.

4 Methods

Various definitions of vulnerability exist across different fields and disciplines, along with methods to estimate it (Hoddinott and Quisumbing, 2010). The term is commonly recognized as an evaluation of factors that contribute to placing households and infrastructure at risk. A prevalent definition in the literature focuses on the **exposure** to poverty, assessing the **likelihood** of a household’s economic condition deteriorating due to a weather event (Baquie and Fuje, 2020; Chaudhuri, 2003; Chaudhuri et al., 2002; Dutta et al., 2011; Günther and Harttgen, 2009; Hill and Porter, 2017; Pritchett et al., 2000; Skoufias and Baez, 2021; Skoufias et al., 2019; Suryahadi and Sumarto, 2003). We use this definition of vulnerability, linking vulnerability directly to poverty, which enables the identification of households at risk of becoming poor in the future, regardless of their current well-being.

We define vulnerability as exceeding a 50% likelihood of falling into poverty in response to weather variation, similar to seminal works such as Baquie and Fuje (2020); Chaudhuri (2003); Skoufias et al. (2019). This definition yields a binary measure of household vulnerability that is insensitive to the depth (intensity) of poverty, is easy to understand and interpret (Hill and Porter, 2017). Similar to prior work on vulnerability by Skoufias et al. (2019) and Hill and Porter (2017), this analysis has as its foundation estimating the mean

and variance of consumption with respect to weather variation and household level characteristics.

The vulnerability analysis can be decomposed into four main steps (1) estimating mean consumption (2) estimating variance of consumption through simulation (3) calculating the probability of falling into poverty (4) classifying household vulnerability based on probabilities from step (3) relative to a common probability benchmark.

Note we assume that (1) household welfare can be measured by the total per capita consumption, defined as the sum of food and non-food consumptions³ (2) household consumption can be determined by combining data on the mean of consumption without weather shocks and the variance of consumption as a function of household characteristics and weather variation. These assumptions enable us to calculate vulnerability, i.e., the probability of each household falling below the international poverty line of \$1.9 US dollars (given in 2011 purchasing power parity) per day consumption. We expect that the vulnerability measure will be mainly driven by the variance of household consumption, i.e., we expect that households with a larger value of variance may be more vulnerable to poverty. However, a high variance may not translate to vulnerability when mean household consumption is sufficiently above the poverty line.

4.1 Estimating vulnerability to poverty

In the following subsections, we present each step of vulnerability to poverty analysis in detail.

4.1.1 Estimating mean consumption

First, we estimate the mean of each household's consumption based on the relationship between the household-specific observable characteristics in the absence of any weather shocks, i.e.,:

$$\ln C_{hdt} = \beta_0 + \beta_1 X_{hdt} + \mu_d + \varepsilon_{hdt} \quad (2)$$

where C_{hdt} denotes the log of per capita household consumption (in 2011 purchasing power parity (PPP) terms, measured in US dollars) of household h at location d in time t , X_{hdt} denotes household-level characteristics (e.g., household size), μ_d stands for the regional fixed effects, ε denotes the error term with the usual properties of zero mean and constant variance, and β 's represent the parameters to be estimated. The

³A common alternative is using income variations. In our setting income data is not readily available.

fitted values of equation 2 represent the expected mean of logged household consumption in the absence of weather shocks, which we can represent as:

$$E(\ln C_{hdt}) = \hat{\beta}_0 + \hat{\beta}_1 X_{hdt} + \hat{\mu}_d \quad (3)$$

4.1.2 Estimating variance from cross-sectional data via simulation

Second, we estimate the variance of future consumption based on household characteristics and indicators of agro-environmental conditions. Ideally, we would like to have repeated observations of the same household's total consumption over time to estimate its variance⁴. However, since we have cross-sectional data, we estimate the variance by examining the spread emerging from 100 draws of the weather distribution and the error term, which yields 100 observations of the total per capita consumption for each household within our sample.

This process involves first estimating the relationship between household characteristics, indicators of agro-environmental conditions, and the log of total per capita consumption per equation 4. In other words, we add a weather term to equation 2:

$$\ln C_{hdt} = \gamma_0 + \gamma_1 X_{hdt} + \gamma_2 W_{dt} + \mu_d + \epsilon_{hdt} \quad (4)$$

where W_{dt} represents the weather indicator in location d and time t , ϵ denotes the error term with usual properties of zero mean and constant variance, and γ 's are the coefficients to be estimated.

To estimate the variance of total consumption, we take draws from the (hypothesized) historical weather distribution as well as the error term from equation 4.

$$E(\ln C_{hdt_{MC}}) = \hat{\gamma}_0 + \hat{\gamma}_1 X_{hdt} + \hat{\gamma}_2 \cdot W_{MC} + \hat{\mu}_d + \epsilon_{MC} \quad (5)$$

where $\hat{\gamma}$ s are the estimated coefficients from equation 4, W_{MC} represent 100 simulated draws from the distribution of weather indicator, and ϵ_{MC} represents 100 simulated draws from the distribution of error term.

⁴In the case of panel data, we can simply take the variance of the repeated observations.

We then combine those simulated values with the coefficients as estimated from equation 4 to yield 100 different estimates of the log of total per capita consumption under different potential states of the world. This approach yields different realizations of the (log-transformed) total consumption corresponding to *future* weather shocks and unobservables. We obtain the standard deviation and variance of household consumption based on the 100 simulated values calculated in equation 5.

4.1.3 Calculating probability of falling into poverty

Third, we estimate the probability of the household falling into poverty using the expected mean and standard deviation of consumption as calculated above. The probability that a household's consumption will fall below the poverty line can be calculated as follows:

$$P[E(\ln C_{hdt+1}) \leq \ln z] = \Phi\left(\frac{\ln z - E(\ln C_{hdt+1})}{\sqrt{\text{var}(\ln C_{hdt+1})}}\right) \quad (6)$$

where z reflects the poverty line (US \$1.9, in 2011 PPP) and $\Phi(\cdot)$ represents the cumulative density function of the standard log-normal distribution.⁵

4.1.4 Determining household vulnerability as a function of its probability of falling into poverty

Finally, we use the probabilities from equation 6 to determine whether the household is considered vulnerable to poverty or not. Specifically, if the probability of falling into poverty is higher than the commonly used 0.5 or 50% threshold, we classify that household as vulnerable to poverty. If the threshold changes, for instance, we lower it to 30%, then more households would be classified as vulnerable to poverty. This might capture more households who are at risk, but this could also lead to more false positive cases, i.e., misclassifying non-vulnerable households. Therefore, the 50% threshold can be considered as a more neutral and easy-to-interpret point of comparison.

4.2 Robustness of vulnerability estimates to changing input assumptions

In this subsection, we highlight the importance of assessing the sensitivity of vulnerability estimates to alterations in input assumptions, such as chosen statistical distributions during the simulation phase for

⁵Examination of the empirical distribution of the consumption data across all years suggests that assuming a log-normal distribution is appropriate.

weather anomaly measures and household-level characteristics.

4.2.1 Examining the impact of historical weather data distribution on vulnerability

Understanding how different distributional assumptions impact vulnerability estimates is crucial for several reasons. Firstly, weather data often exhibit diverse characteristics that might not be adequately captured by a single distributional assumption. For instance, while some variables like rainfall may follow a skewed distribution with positive values only, others such as temperature may exhibit symmetric distributions with both positive and negative values. Failing to account for these nuances can distort vulnerability estimates and undermine their reliability.

Furthermore, by considering alternative statistical distributions commonly used in weather literature, such as Log-normal, Exponential, Gamma, Weibull, and Logistic distributions, we aim to provide a more comprehensive analysis of how different assumptions affect vulnerability estimates. Each distribution is selected based on its suitability for modeling specific agro-environmental variables. For example, Log-normal, Exponential, Gamma, and Weibull distributions are well-suited for positively defined variables, such as rainfall data, while Logistic distribution is commonly used to characterize symmetrically distributed variables and those with heavy tails, such as NDVI and extreme precipitation events.

By exploring the sensitivity of vulnerability estimates to different distributional assumptions, we can enhance the robustness and reliability of vulnerability assessments. This comprehensive analysis enables us to better understand the potential implications of different assumptions on vulnerability estimates and make more informed decisions in climate risk management and adaptation planning. In the following subsections, we provide an overview of the tools we use for this purpose. We then proceed to select one of the appropriate alternative distributions and re-estimate our vulnerability analyses, comparing the outcomes to those derived from the analyses using the Gaussian normal distribution.

Cullen and Frey plot

We use the Cullen and Frey plot for the initial comparison of various distribution fits to our selected weather condition measures. This plot employs higher moments of statistical distributions, specifically skewness and kurtosis, to facilitate comparison. Skewness measures the symmetry or relative tail sizes of distributions, while kurtosis gauges the combined weight of the tails relative to the rest of the distribution. A perfectly

symmetric normal distribution has a skewness of 0 and a kurtosis of 3.

The candidate theoretical distributions included in the Cullen and Frey graph for comparison are Normal, Uniform, Exponential, Logistic, Beta, Log-normal, Gamma, and Weibull. In addition to the data and theoretical distributions, the plot displays bootstrapped values, offering insight into the sampling distribution and aiding in the determination of the appropriate distribution for the specific weather anomaly measure. In our example, we employ 50,000 bootstrapped values.

Probability distribution plots

Another visual test for the goodness of fit is probability plots such as Q-Q plots (quantile-quantile plots) and P-P plots (probability-probability plots), each offering distinct insights.

Q-Q plots facilitate a comparison between the quantiles of two datasets. By plotting quantiles of one dataset against quantiles of another, they reveal whether the datasets follow a similar distribution. When the points fall approximately along a straight line, it suggests similarity in distribution between the datasets, aiding in the assessment of model fit or identifying departures from assumed distributions.

P-P plots, on the other hand, compare the cumulative distribution functions (CDFs) of two datasets. They plot the empirical CDF of one dataset against the theoretical CDF of another, providing a visual indication of how well the empirical data matches a theoretical distribution. Similar to Q-Q plots, a straight line in a P-P plot suggests that the datasets come from the same distribution.

Both Q-Q and P-P plots assist in deciding which distribution best fits the data. This decision-making process involves comparing the observed empirical CDF to the theoretical CDFs of various candidate distributions. The distribution that aligns most closely with the empirical CDF, as evidenced by the pattern observed in the Q-Q or P-P plot, is typically chosen as the most appropriate model for the data at hand.

Goodness of fit statistics

To formally compare the fit of a hypothesized distribution to the actual data, we utilize three commonly used goodness-of-fit statistics: the Kolmogorov-Smirnov statistic, the Cramér-von Mises statistic, and the Anderson-Darling statistic.

The **Kolmogorov-Smirnov** (KS) statistic is a widely used method to evaluate the similarity between

empirical and theoretical distributions. It calculates the maximum vertical distance between the empirical cumulative distribution function (CDF) derived from the sample data and the theoretical CDF. This distance, known as the KS statistic, serves as a measure of the largest discrepancy between the observed and expected cumulative probabilities. A smaller KS statistic suggests a closer match between the empirical and theoretical distributions, indicating a better fit. The KS test is non-parametric, meaning it does not rely on specific assumptions about the underlying distribution, making it versatile for various types of data. However, it tends to be more sensitive to differences in the central part of the distribution and less sensitive to differences in the tails.

The **Cramer-von Mises** (CvM) statistic offers an alternative approach to assess the goodness of fit between observed and theoretical distributions. Unlike the KS test, which considers only the maximum deviation between the empirical and theoretical CDFs, the CvM statistic evaluates the overall discrepancy by summing the squared differences between corresponding CDF values. This provides a more balanced assessment across the entire distribution, giving greater weight to discrepancies in the tails. Consequently, the CvM test is often more sensitive to differences in the tails of the distribution compared to the KS test. However, it still maintains the advantage of being distribution-free, making it applicable to a wide range of data types without requiring assumptions about the underlying distribution. Overall, the CvM test provides a robust measure of goodness of fit that considers both central tendency and tail behavior.

The **Anderson-Darling** (AD) statistic is another method commonly used for assessing distributional fit. It builds upon the concepts of the KS and CvM tests by placing greater emphasis on differences in the tails of the distribution. The AD statistic incorporates a weighting function that assigns a higher weight to discrepancies in the tails, making it particularly sensitive to deviations in these regions. This makes the AD test especially useful for detecting differences in extreme values, which may be critical in certain applications such as risk assessment or reliability analysis. However, like the other two tests, the AD test is also non-parametric and does not rely on specific distributional assumptions. This flexibility allows it to be applied to a wide range of data types

4.2.2 Assessing additional household characteristics as potential mediators

Finally, to examine the mediating influence of household-level characteristics on vulnerability, we conduct the same vulnerability analysis but include different household-level characteristics as controls. We use the list of household characteristics that are widely used in the prior literature and are hypothesized to mediate

the relationship between weather anomalies and vulnerability to poverty measures (Baquie and Fuje, 2020; Chaudhuri, 2003; Hill and Porter, 2017; Skoufias et al., 2019).

The baseline analysis considers only household size. Then, we sequentially add additional characteristics including characteristics of the household head (i.e., gender, age, education level, marital status), asset ownership (e.g., livestock, car), the type of residential building materials used for the walls, roof, and floor of the household’s residence. Keeping all the other factors constant, we compute the average vulnerability estimate for each model, employing it as a basis for comparison. We anticipate that household size will have a negative impact on the total per capita consumption, while higher education levels and better living conditions will have a positive effect on the total consumption.

5 Results and Discussion

This section provides a summary of the primary findings derived from vulnerability analyses conducted on rural households in Niger. Our results indicate that estimates of vulnerability to poverty are influenced by the household-level characteristics included in the analysis. However, the influence of the statistical distribution on the vulnerability estimate varies. While the choice of statistical distribution has little effect on the vulnerability estimate for rainfall, it significantly influences the estimate for the analysis using the vegetation index measure.

5.1 Vulnerability Measure

Figure 12 summarizes the results of vulnerability analysis from approximately 20,500⁶ estimated models. The graphic illustrates that $\sim 73\%$ of these models, equivalent to around 15,000 instances out of the total 20,500 models, exhibit vulnerability estimates within the range of 31.5% to 32%. Nonetheless, there are instances where the vulnerability estimate rises as high as 40%. To gain a deeper understanding of the factors influencing such variations in vulnerability estimates, we present average vulnerability estimates obtained from various model specifications.

Figure 13 displays average vulnerability estimates derived from an analysis employing cumulative rainfall, seasonal rainfall, NDVI, and standardized measures of rainfall and vegetation index as proxies for agro-

⁶We have $2^{11} = 2048$ combinations of household-level characteristics. For each combination, we evaluate 5 weather anomaly measures, and each anomaly measure is assessed using 2 statistical distributions. Hence, in total, we estimate $2048 \times 5 \times 2 = 20,480$ models.

environmental conditions. The separate results for rainfall and NDVI weather condition indicators are presented in Figures 14 - 17. For ease of visualization, we present only 50 randomly selected model outputs from the pool of over 20,000 models in a specification chart.

In the specifications chart, each data point represents the average vulnerability estimate of the model. The lower portion of Figure 13 provides specific details regarding the model, including the weather anomaly measure, statistical distribution, and household-level characteristics considered. For example, the blue dot in Figure 13 corresponds to an average vulnerability estimate of approximately 31.8%. This estimate comes from an estimated model that uses the *NDVI* measure evaluated with log-normal statistical distribution. Moreover, the model incorporates household-level variables such as *household size*, *the primary material for wall construction*, *access to sanitation facilities*, and *personal car ownership status*.

We find that vulnerability estimates span from 31% to 41%. A significant discrepancy of 10 percentage points, equivalent to roughly 2 million people, is primarily attributed to the selection of specific weather condition metrics. Particularly, we notice that estimates tend to be higher when using the seasonal precipitation measure. This observation broadly aligns with our expectations.

The rationale behind this observation lies in the nature of the seasonal precipitation variable, which is derived from primary growing season rainfall data. This approach ensures that the seasonal rainfall variable reflects the conditions during the growing season in Niger. In contrast, cumulative precipitation measure factors in data from all months, which could potentially distort the representation of rainfall variation, especially when rainfall during non-growing months remains consistently low. Hence, we can assume that the seasonal precipitation variable may be better suited to capture consumption variation due to its direct alignment with the timing of the growing season since household consumption patterns are closely tied to crop availability, primarily determined by rainfall during the crucial growing period.

Additionally, we examine how vulnerability estimates vary across Niger to determine if the distribution of vulnerable households remains consistent across different models. Figure 18 represents the distribution of vulnerable households across time and space for seasonal rainfall measure ⁷.

Figure 18 illustrates the highest concentration of vulnerable households in the southern and south-western regions of Niger. This spatial distribution is in line with our expectations, as these areas heavily rely on rainfed agricultural activities. Notably, the southern administrative units of Tillaberi, Dosso, Maradi,

⁷The distribution for the remaining weather condition measures is consistent with the one presented in Figure 18.

Diffa, and Tahoua stand out as the most vulnerable. These regions collectively contribute to 90% of the country’s millet production—a vital rainfed agricultural crop for the country ([International Production Assessment Division, 2021](#)). The prominence of millet production underscores the critical role of these regions in the country’s agricultural landscape. Moreover, it suggests that households in these areas are particularly susceptible to adverse weather conditions, significantly increasing their risk of slipping into poverty.

The temporal distribution of vulnerability estimates, as depicted in [Figure 18](#), indicates a peak in the severity of the results in 2011 compared to 2014 and 2018. This observed pattern aligns with the relatively adverse weather conditions during that period, as illustrated in [Figure 3](#) and discussed in [Section 3](#). This finding solidifies the notion that anomalous weather events have a significant and negative impact on household welfare. For decision-makers, it underscores the critical importance of integrating consumption-weather analysis into the development and implementation of climate-resilient strategies to effectively mitigate these weather-related effects.

5.2 Sensitivity of vulnerability estimates to changing input assumptions

In this section, we present the results of sensitivity analyses of vulnerability estimates. Our findings indicate that the vulnerability estimate is sensitive to the choice of household-level characteristics; whereas the choice of statistical distribution suitable for each weather anomaly measure leads to varying results. Additionally, we discuss how decision-makers can leverage our results to inform their social protection programs.

5.2.1 Results for choice of historical distribution

In this section, we outline the steps we undertook to identify the suitable alternative statistical distribution for each of the weather anomaly measures. We then proceed to select one of these distributions and re-estimate our vulnerability analyses, comparing the outcomes to those derived from the Gaussian normal distribution.

We start by looking at the empirical densities of weather condition measures. [Figure 6](#) presents histograms of anomaly measures. Besides the standardized measure of rainfall, none of the other measures closely resemble a normal distribution.

When we examine the data more closely, we notice certain trends. For instance, mean NDVI and rainfall

measures are right skewed, which is what we would expect. Rainfall data typically exhibits a right-skewed distribution due to the sporadic occurrence of heavy rainfall events, which disproportionately contribute to the overall precipitation. Similarly, the right skewness observed in mean NDVI is consistent with the fact that higher vegetation levels are generally less prevalent, resulting in a distribution skewed towards lower values.

On the other hand, the left skewness observed in the standardized measure of NDVI may seem unexpected at first glance. However, this can be attributed to the normalization process used. Negative deviations from the mean NDVI could be causing the observed left skew in the distribution. These observations highlight the complex nature of environmental data and importance of carefully selecting appropriate statistical distributions for each agro-environmental condition measure.

Cullen and Frey plot

In Figure 7, we have a Cullen and Frey graph, which is also known as a skewness-kurtosis plot, for all selected weather condition measures. It illustrates the results of our analysis on potential statistical distribution fits for the selected weather condition measures. We observe that for cumulative rainfall, distributions such as Gamma, Log-normal, and Weibull might offer a more suitable fit. Conversely, for NDVI, the Logistic distribution appears to be appropriate. Additionally, when examining the standardized measure of rainfall, we notice that the observed values and bootstrapped results are closely located to Logistic and Log-normal distributions. While for the standardized measure of NDVI, Log-normal and Gamma distributions might provide a better fit.

Probability distribution plots and goodness of fit statistics

Figures 8 - 11 provide additional insight into the goodness of fit of various distributions to weather condition data. Each figure presents histograms, Q-Q plots, P-P plots, and empirical and theoretical cumulative distributions.

The analysis reveals that for the rainfall data, the Exponential, Gamma, and Weibull distributions closely match the observed data, suggesting their appropriateness for modeling rainfall patterns. Conversely, for the NDVI data, all distributions except the Exponential distribution demonstrate a close fit to the observed values, indicating their effectiveness in capturing NDVI measures' distribution. At the same time, the

Logistic distribution exhibits superior fit to both standardized rainfall and NDVI measures compared to other distributions.

Goodness of fit statistics

The above-discussed visual tests do not demonstrate whether the fit is statistically different from the baseline of normal distribution. Hence, we also use goodness-of-fit statistics such as Kolmogorov-Smirnov (KS), Cramer-von Mises (CvM), and Anderson-Darling (AD) to evaluate whether the alternative distributions produce a statistically significantly better fit to our data. Table 4 summarizes the results for both visual tests and the goodness-of-fit statistic. Based on our analyses, we choose Weibull distribution for cumulative rainfall, Log-normal for NDVI, and logistic distribution for standardized measures as an alternative to the benchmark Normal distribution as they produce a statistically significantly better fit for our weather data.

After selecting an appropriate distribution for each weather condition measure, we rerun our analysis using these distributions. We later compare the results from the alternative distributions to those of the Gaussian Normal distribution.

We find that switching from a Normal distribution to Weibull for cumulative rainfall brings a marginal difference of 0.008% (equivalent to 1,625 households) in the vulnerability estimate. Similarly, moving from Normal to Logistic distribution for standardized rainfall measures leads to a 0.15% difference (equivalent to 30,480 households) in the vulnerability estimate. This finding suggests that while the alternative distributions may indeed capture the variation in the underlying data better, the slight difference in the final vulnerability estimate implies that researchers might find it simpler to stick to a Normal distribution when using rainfall data as a proxy for weather conditions, rather than complicating the estimation framework.⁸

The results are more pronounced for the vegetation index. Specifically, shifting from a Normal to a Log-normal distribution for the NDVI measure produces a 0.3% difference in vulnerability estimates, affecting approximately 61,000 Nigerien households. Conversely, using a Logistic distribution instead of a Normal distribution for the standardized NDVI measure results in a 0.74% change in the vulnerability estimate, impacting around 150,000 additional households. Comparing these findings to those of rainfall, it is evident that the choice of statistical distribution has a more significant impact on results in the case of the vegetation index. This is likely due to the nature of the vegetation index, which reflects the health of vegetation in the

⁸We have determined that the appropriate alternative distribution for seasonal precipitation is a Normal distribution. Consequently, there is no ongoing discussion regarding differences in results concerning seasonal rainfall data.

area.

These results may not always be generalizable. However, given the sensitivity to the choice of statistical distribution observed here, we expect to find larger deviations in results when using the dataset covering more than three years. Hence, we advise policymakers to regularly check the distributional assumptions for the weather condition measures (especially, vegetation index) and select the one that fits the data closely.

5.2.2 Results for additional household characteristics included in the analysis

In examining the vulnerability results for additional household characteristics, we use the regression analysis with agro-environmental variables and household size as the benchmark. We find that vulnerability varies, but slightly, when we account for a single additional household characteristic. However, when holding all the other components constant, we find that compared to the model that accounts only for household size, the inclusion of almost all household-level characteristics leads to 6% point difference in vulnerability estimates (equivalent to 1.2 million households being misidentified as non-vulnerable).

Further analyses show that the key drivers of vulnerability estimate are household size, household head age and sex, household living conditions, and ownership of car variables. Therefore, we suggest that practitioners and policymakers conducting vulnerability analyses include household-level characteristics that not only significantly influence vulnerability estimates but also have been extensively utilized in previous research and are among the most commonly employed measures.

This approach can be especially beneficial in situations where detailed household-level information is scarce (which is the case in many developing countries). When faced with such constraints, decision-makers can opt to prioritize gathering core household characteristics that have a substantial impact on vulnerability estimates. By conducting concise surveys focused on collecting these key characteristics, not only can survey efforts be simplified for greater efficiency, but they can also prove to be more cost-effective. This offers decision-makers a practical and resource-efficient strategy for acquiring essential information for conducting vulnerability analysis.

6 Conclusion

This paper estimates the rural population's vulnerability to poverty in Niger and analyzes potential underlying drivers. Our findings indicate that vulnerability estimates range from 31% to 41%, implying that

approximately 6.5 to 8.3 million households face an increased risk of slipping into poverty due to unfavorable weather events. Through a systematic evaluation of key factors, commonly encountered by scholars and practitioners, we find that vulnerability estimates are sensitive to the choice of household-level characteristics. We identify *household size, household head age and sex, household living conditions, and ownership of car* variables as the main drivers of vulnerability estimate. This finding is important for decision-makers, especially when conducting a vulnerability estimate for developing countries lacking comprehensive household livelihood surveys. In these instances, this finding can aid policymakers in identifying and gathering only essential household-level variables. This targeted approach can be more cost-effective compared to conducting extensive household surveys. It offers a quick and direct means of gathering crucial information for conducting vulnerability analysis essential for designing flexible and adaptive social protection programs.

We also find that the choice of appropriate statistical distribution leads to varying results depending on the weather condition measure considered. Specifically, our analysis reveals that vulnerability estimates derived from models incorporating rainfall data remain consistent across different statistical distributions. However, when focusing on the vegetation index, we observe significant variability in the results depending on the chosen statistical distribution. This highlights the importance of choosing a statistical distribution that accurately represents the data as a well-fitted distribution provides valuable insights into the data, aiding in better understanding and decision-making. Thus, the importance of selecting the right statistical distribution cannot be overstated, as it directly influences the vulnerability estimate.

Overall, the insights derived from our analysis not only highlight the potential severity of future weather-related challenges in Niger but also serve as a valuable tool for policymakers in developing effective social protection programs and projects. By understanding the distinctive features of household vulnerabilities, policymakers can tailor interventions that address the specific characteristics of the impending crisis.

Bibliography

- Arouri, M., Nguyen, C., Youssef, A.B., 2015. Natural disasters, household welfare, and resilience: evidence from rural vietnam. *World development* 70, 59–77.
- Baqaie, S., Fuje, H., 2020. Vulnerability to Poverty following Extreme Weather Events in Malawi.
- Brunelin, S., Amadou, M.M., Lung, F., Plevko, S., 2022. Responding faster to droughts with satellites and adaptive social protection in niger.
- Central Intelligence Agency, 2024. *The world factbook: Niger*.
- Chaudhuri, S., 2000. Empirical methods for assessing household vulnerability to poverty. School of International and Public Affairs, Columbia University, New York .
- Chaudhuri, S., 2003. Assessing vulnerability to poverty: concepts, empirical methods and illustrative examples. Department of Economics, Columbia University, New York 56.
- Chaudhuri, S., Jalan, J., Suryahadi, A., 2002. Assessing household vulnerability to poverty from cross-sectional data: A methodology and estimates from indonesia .
- Christiaensen, L.J., Sarris, A., 2007. Rural household vulnerability and insurance against commodity risks: Evidence from the United Republic of Tanzania. volume 10. Food & Agriculture Organization.
- Climates to Travel, 2023. *Climate - niger*.
- Concern Worldwide U.S., 2022. *Poverty in niger: 5 things you need to know*.
- Dercon, S., Hoddinott, J., Woldehanna, T., 2005. Shocks and consumption in 15 ethiopian villages, 1999–2004. *Journal of African economies* 14, 559–585.
- Dutta, I., Foster, J., Mishra, A., 2011. On measuring vulnerability to poverty. *Social Choice and Welfare* 37, 743.
- Food & Agriculture Organization, 2019. *The State of Food Security and Nutrition in the World: Safeguarding Against Economic Slowdowns and Downturns*. Technical Report. Food & Agriculture Organization.
- Funk, C., Peterson, P., Landsfeld, M., Pedreros, D., Verdin, J., Shukla, S., Husak, G., Rowland, J., Harrison, L., Hoell, A., et al., 2015. The climate hazards infrared precipitation with stations—a new environmental record for monitoring extremes. *Scientific data* 2, 1–21.
- Funk, C.C., Rowland, J., Eilerts, G., Adoum, A., White, L., 2012. A climate trend analysis of niger.
- Gao, J., Mills, B.F., 2018. Weather shocks, coping strategies, and consumption dynamics in rural ethiopia. *World Development* 101, 268–283.

- Günther, I., Harttgen, K., 2009. Estimating households vulnerability to idiosyncratic and covariate shocks: A novel method applied in madagascar. *World Development* 37, 1222–1234.
- Hallegatte, S., Bangalore, M., Bonzanigo, L., Fay, M., Kane, T., Narloch, U., Rozenberg, J., Treguer, D., Vogt-Schilb, A., 2016. *Shock Waves: Managing the Impacts of Climate Change on Poverty*. Climate Change and Development Series. Washington DC: World Bank.
- Heinrigs, P., 2010. *Security implications of climate change in the sahel region: policy considerations*. Paris: Organization for Economic Cooperation and Development Sahel and West Africa Club Secretariat .
- Held, I., 2000. *Sahel drought: Understanding the past and projecting into the future*. NOAA .
- Hill, R.V., Porter, C., 2017. Vulnerability to drought and food price shocks: evidence from ethiopia. *World Development* 96, 65–77.
- Hirvonen, K., 2016. Temperature changes, household consumption, and internal migration: evidence from tanzania. *American Journal of Agricultural Economics* 98, 1230–1249.
- Hoddinott, J., 2006. Shocks and their consequences across and within household in rural zimbabwe.[www document] url: <http://www.sarpn.za/document>.
- Hoddinott, J., Quisumbing, A., 2010. *Methods for microeconomic risk and vulnerability assessment*. Springer.
- International Production Assessment Division, 2021. *Niger millet area, yield, and production*.
- Kurosaki, T., 2015. Vulnerability of household consumption to floods and droughts in developing countries: evidence from pakistan. *Environment and Development Economics* 20, 209–235.
- Lohmann, S., Lechtenfeld, T., 2015. The effect of drought on health outcomes and health expenditures in rural vietnam. *World development* 72, 432–448.
- Lu, Y., 2019. *Sahel climate region*. ESRI .
- Mahamadou, O., Murthi, M., 2022. *How niger’s safety net helps its most vulnerable citizens thrive amid crises*. Atlantic Council .
- Mera, G.A., 2018. Drought and its impacts in ethiopia. *Weather and climate extremes* 22, 24–35.
- Mirza, M.M.Q., 2003. Climate change and extreme weather events: can developing countries adapt? *Climate policy* 3, 233–248.
- Mortimore, M., 2010. Adapting to drought in the sahel: lessons for climate change. *Wiley interdisciplinary reviews: climate change* 1, 134–143.

- Nguyen, G., Nguyen, T.T., 2020. Exposure to weather shocks: A comparison between self-reported record and extreme weather data. *Economic Analysis and Policy* 65, 117–138.
- Population Reference Bureau, 2015. *Demographic Challenges of the Sahel*.
- Porter, C., 2012. Shocks, consumption and income diversification in rural ethiopia. *Journal of Development Studies* 48, 1209–1222.
- Pritchett, L., Suryahadi, A., Sumarto, S., 2000. Quantifying vulnerability to poverty: A proposed measure, with application to indonesia .
- Reynaud, A., Nguyen, M.H., Aubert, C., 2018. Is there a demand for flood insurance in vietnam? results from a choice experiment. *Environmental Economics and Policy Studies* 20, 593–617.
- Sen, A., 1982. *Poverty and famines: an essay on entitlement and deprivation*. Oxford university press.
- Sisa, M., 2014. *Niger food security brief* .
- Skoufias, E., Baez, J., 2021. Operationalizing vulnerability to poverty. *Poverty & Equity Global Practice* .
- Skoufias, E., Kawasoe, Y., Strobl, E., Acosta, P.A., 2019. Identifying the vulnerable to poverty from natural disasters: the case of typhoons in the philippines. *World Bank Policy Research Working Paper* .
- Skoufias, E., Vinha, K., 2012. Climate variability and child height in rural mexico. *Economics & Human Biology* 10, 54–73.
- Suryahadi, A., Sumarto, S., 2003. Poverty and vulnerability in indonesia before and after the economic crisis. *Asian Economic Journal* 17, 45–64.
- Thomas, T., Christiaensen, L., Do, Q.T., Trung, L.D., 2010. Natural disasters and household welfare: evidence from vietnam. *World Bank Policy Research Working Paper* .
- Twongyirwe, R., Mfitumukiza, D., Barasa, B., Naggayi, B.R., Odongo, H., Nyakato, V., Mutoni, G., 2019. Perceived effects of drought on household food security in south-western uganda: Coping responses and determinants. *Weather and Climate Extremes* 24, 100201.
- United Nations, 2019. *Why Population Growth Matters for Sustainable Development*. Technical Report. Department of Economic and Social Affairs at United Nations.
- Weather Spark, 2023. *Climate and average weather year round in niamey, niger*.
- Weier, J., Herring, D., 2000. Measuring vegetation (ndvi & evi) .
- World Bank, 2020a. *Living standards measurement study*.

World Bank, 2020b. Sahel Adaptive Social Protection Program (ASPP).

World Bank Group Climate Change Knowledge Portal, 2021. Niger: Climatology.

World Nuclear Association, 2023. Uranium in niger.

Figures

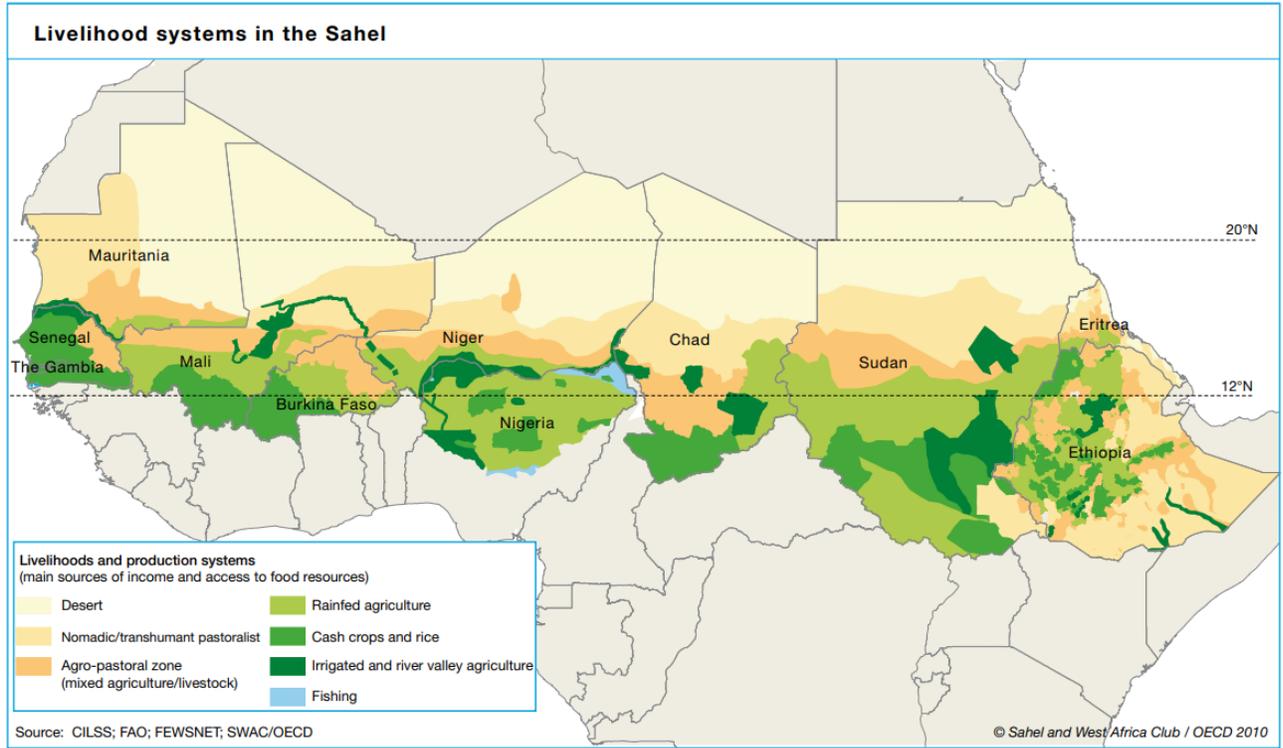


Figure 1: Livelihood Systems in the Sahel

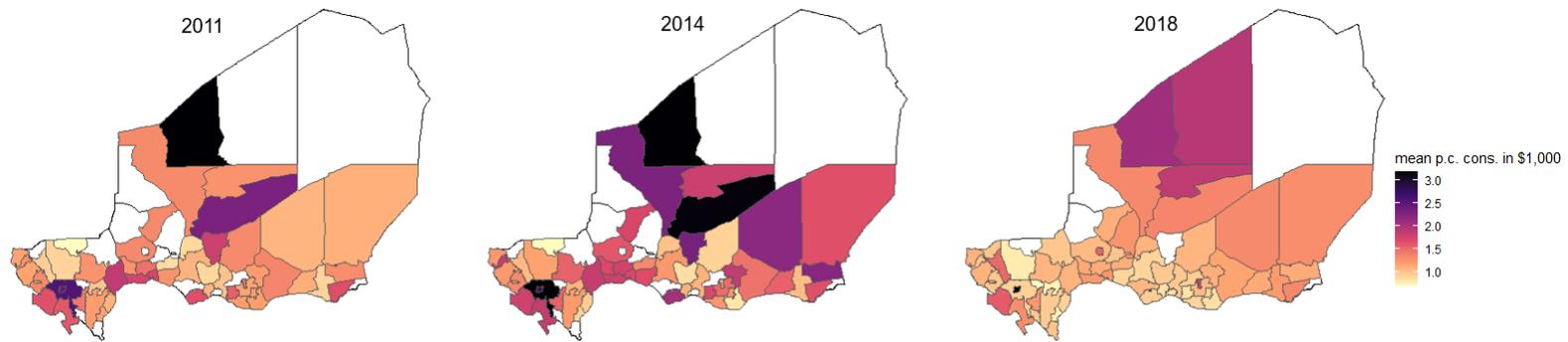


Figure 2: Average annual per capita consumption (in dollars) per each level 2 administrative unit (department) expressed in 2011 prices. Across all years, the north-western and south-western (around the capital city) portions of Niger tend to have higher average p.c. consumption.

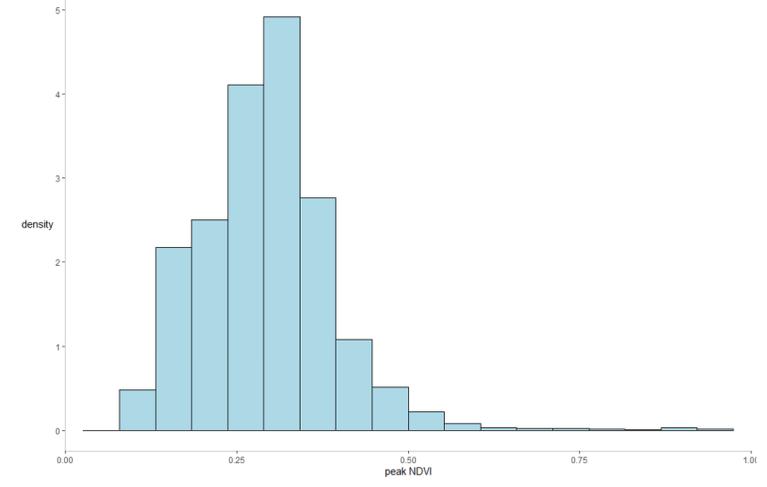
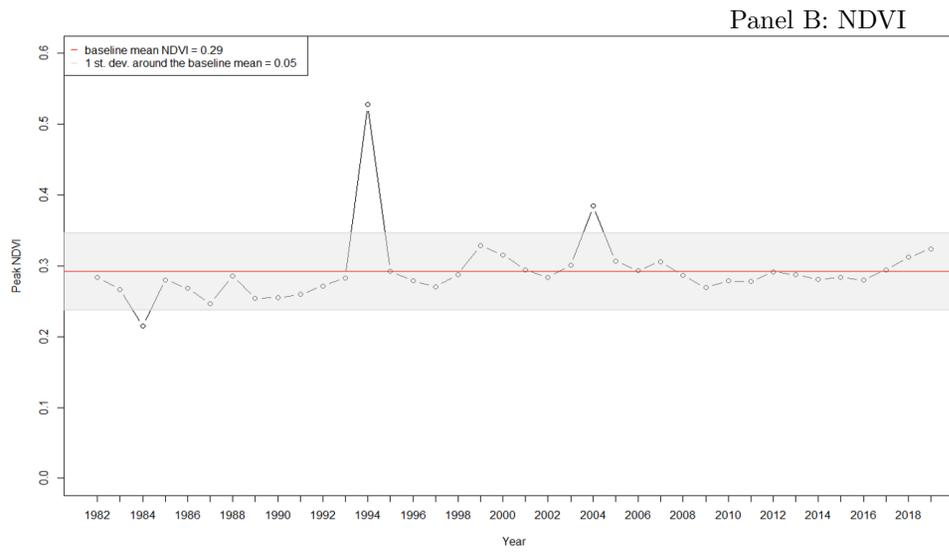
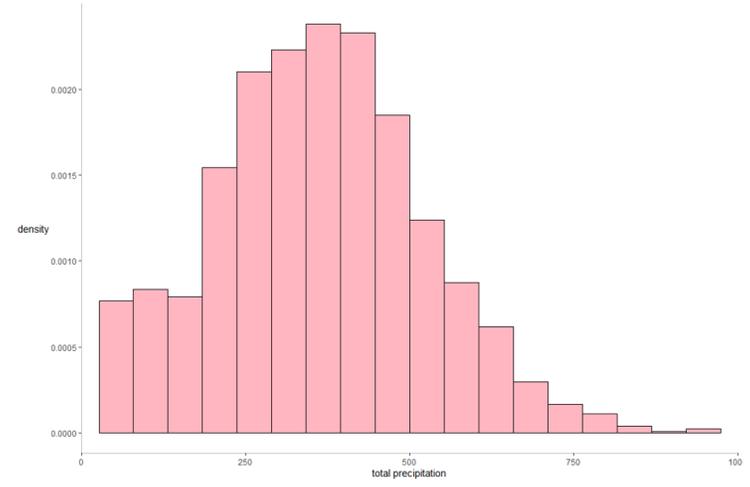
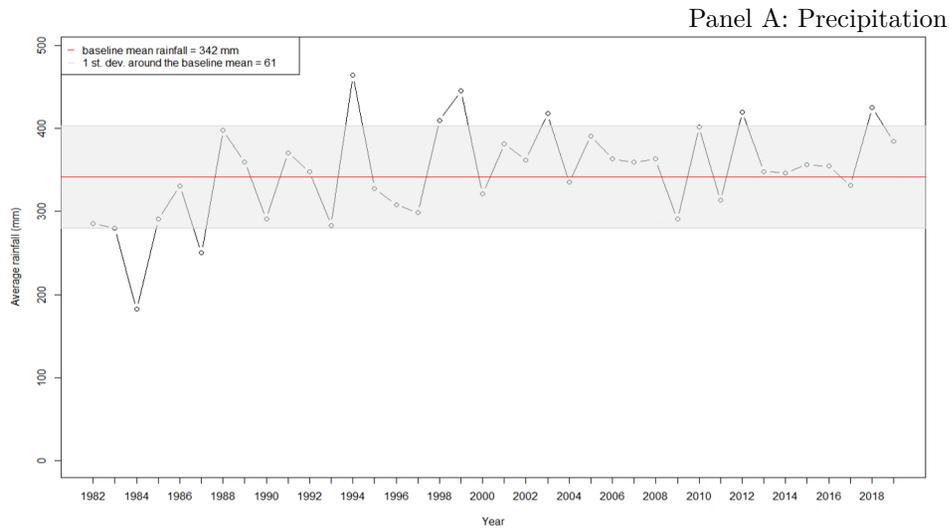
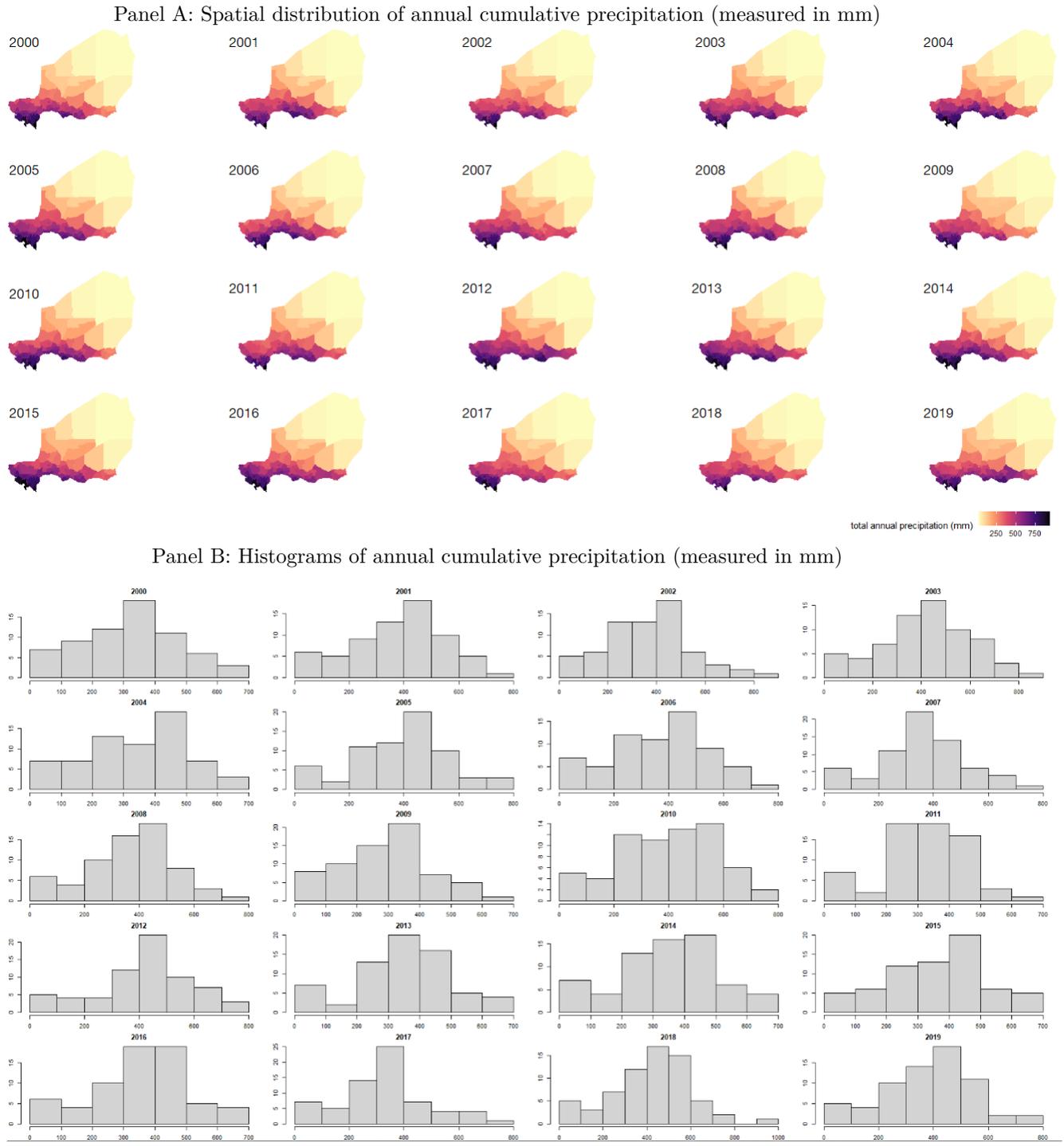


Figure 3: Average precipitation and NDVI from 1982-2019, mean aggregated over the department.



Note: In histograms, the y-axis represents the frequency and the x-axis shows the cumulative precipitation.

Figure 4: Distribution of annual cumulative precipitation (mm) in Niger for the period of 2000 - 2019. Overall, there is not much variation in cumulative precipitation across time. However, within a country across all specified years, the northern regions of Niger receive lower precipitation (due to being close to the Sahara desert) compared to the southwestern regions (due to being close to the savannas).

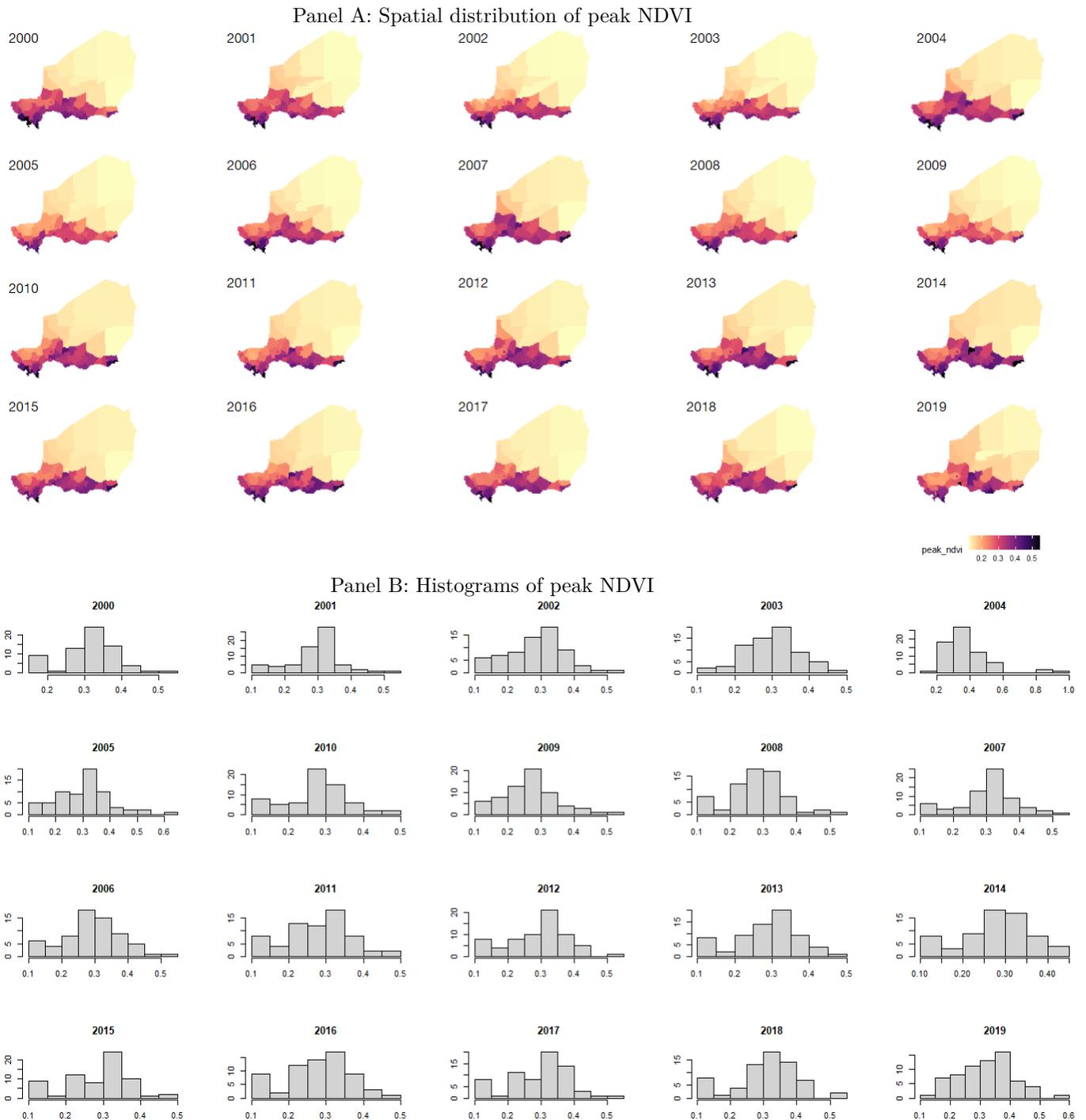


Figure 5: Distribution of peak NDVI in Niger for the period of 2000 - 2019. Overall, the northern regions of Niger record lower NDVI measures compared to the ones in the southern regions. This pattern is expected considering that the northern part of the country neighbors a desert, while the southern regions neighbors the savannas.

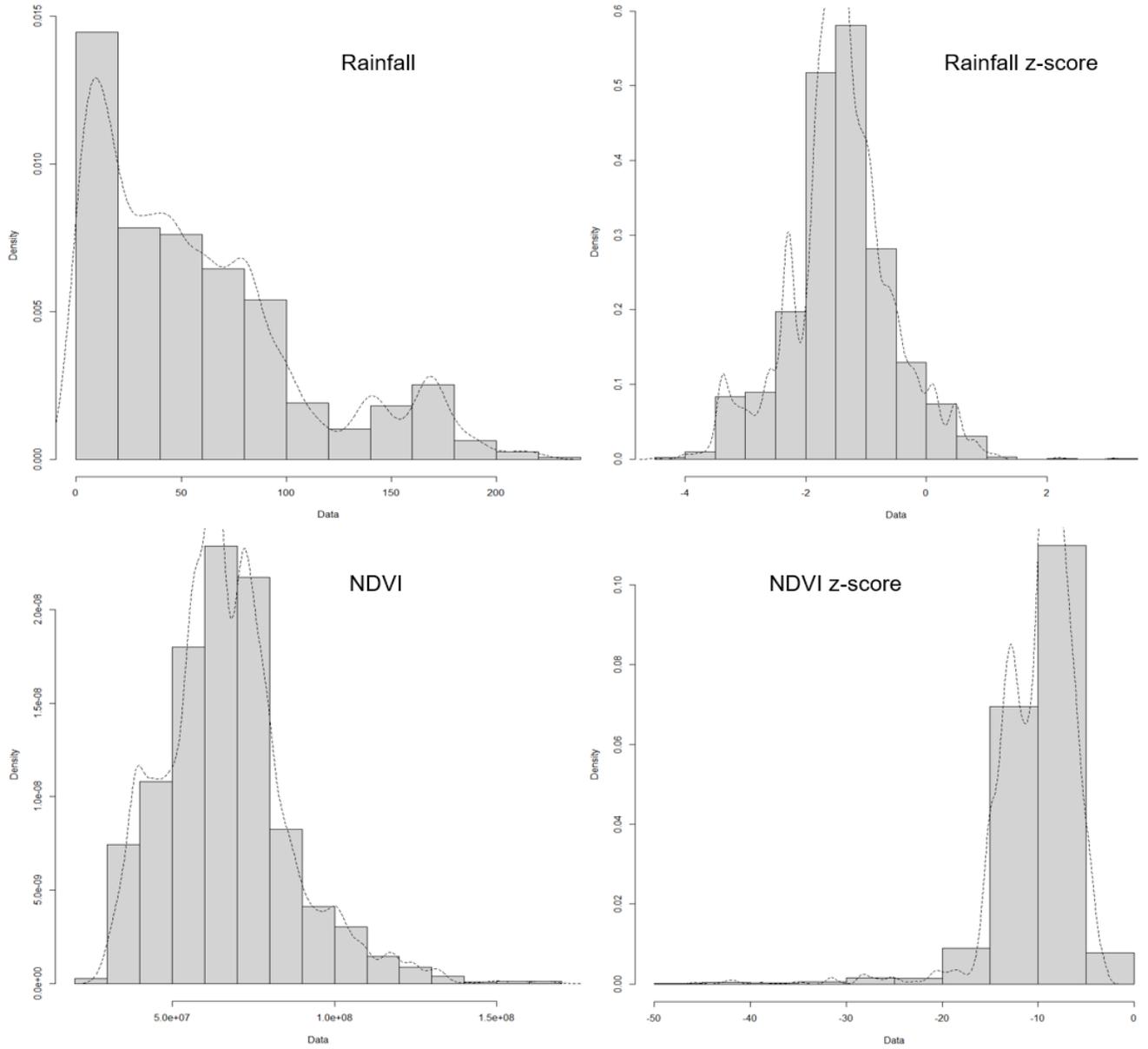


Figure 6: Empirical densities of all selected weather condition measures from the combined dataset. The histograms illustrate that the anomaly measures deviate from normal distribution.

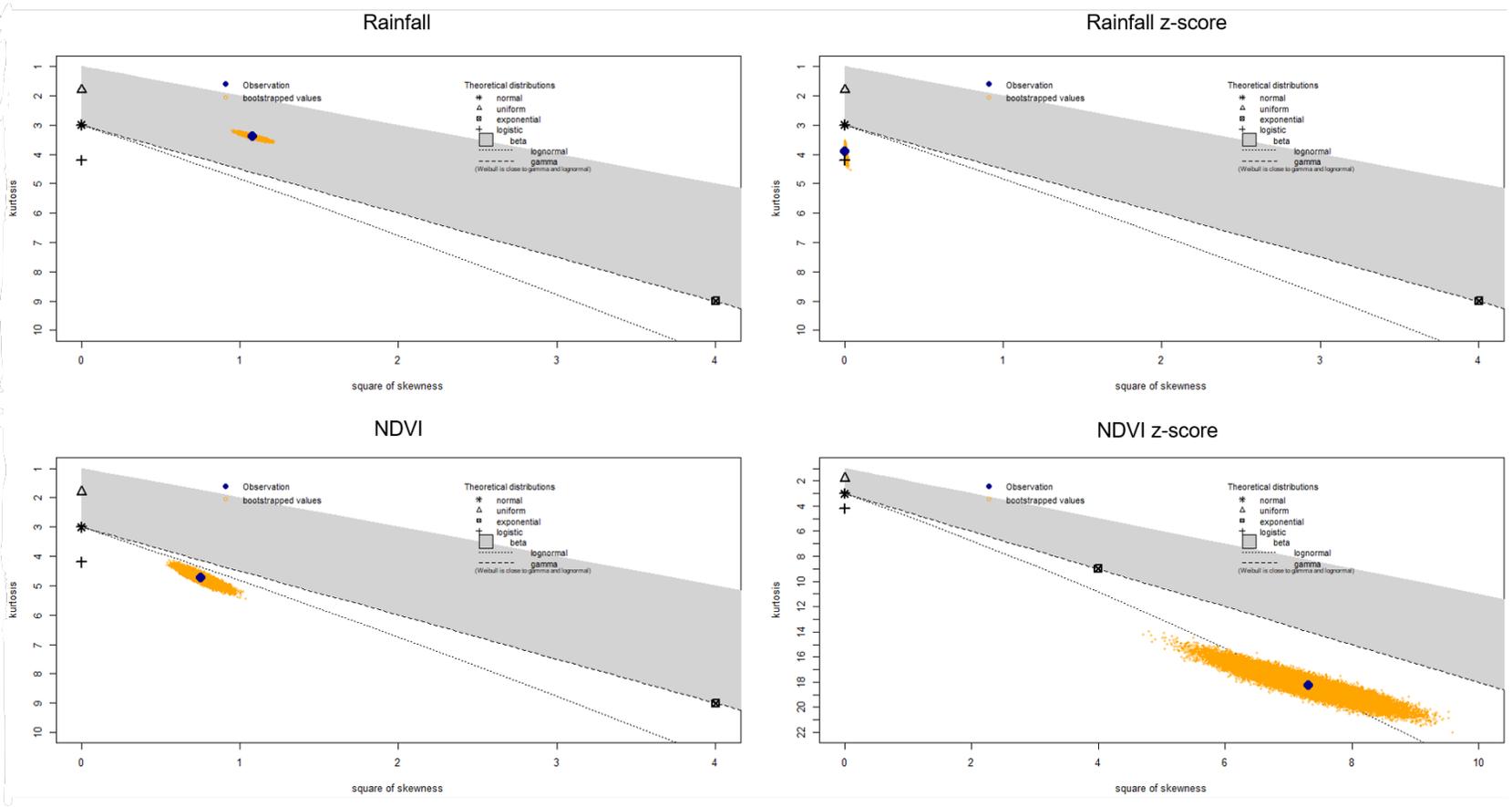


Figure 7: Cullen and Frey graph for all selected weather anomaly measures. It shows that the logistic distribution fits well for NDVI and its standardized measures. However, for rainfall, the potential candidates for distribution are Gamma and Log-normal. A logistic distribution might be more suitable for the standardized measure of rainfall.

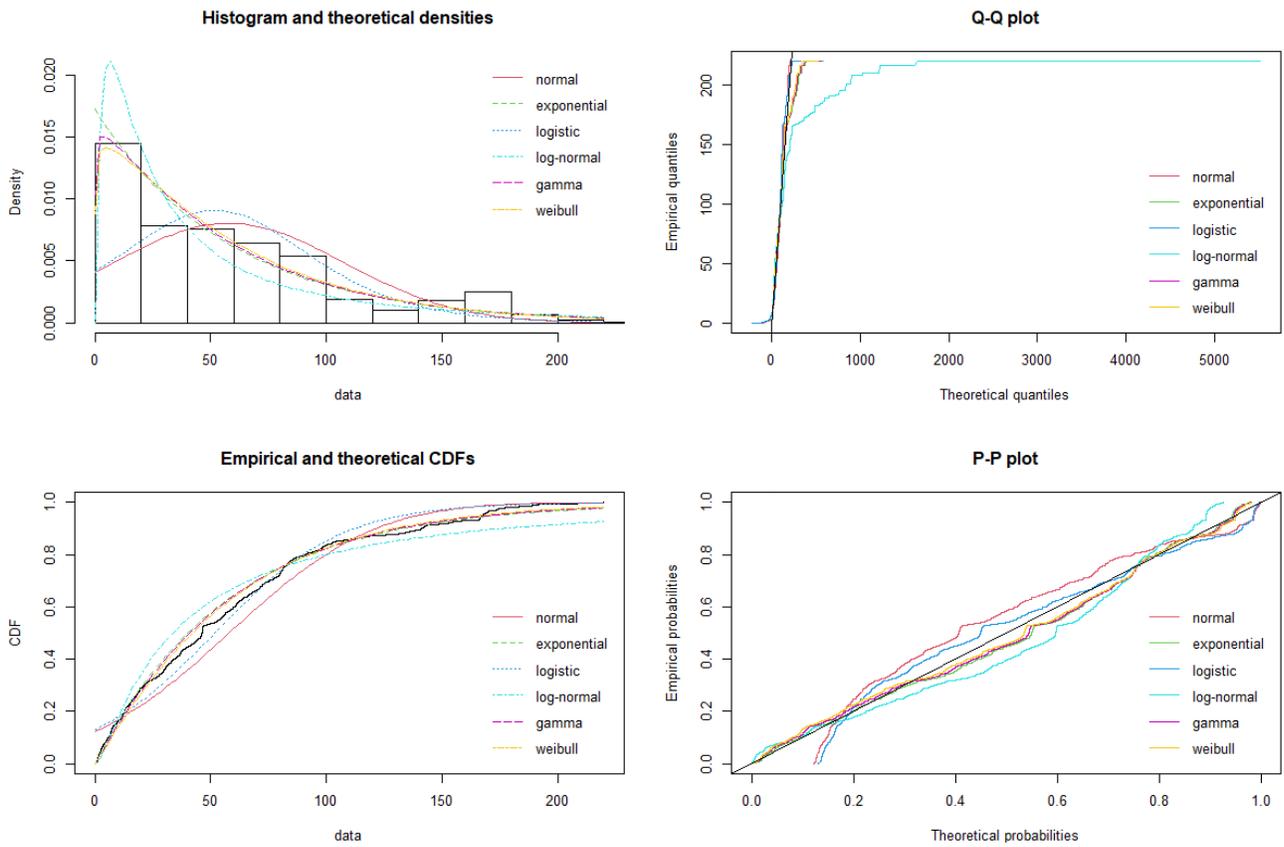


Figure 8: Integrated visual analysis of **cumulative rainfall**: distribution comparison and probability plots. This figure shows that Exponential, Weibull, and Gamma distributions provide a comparatively better fit for the cumulative rainfall data.

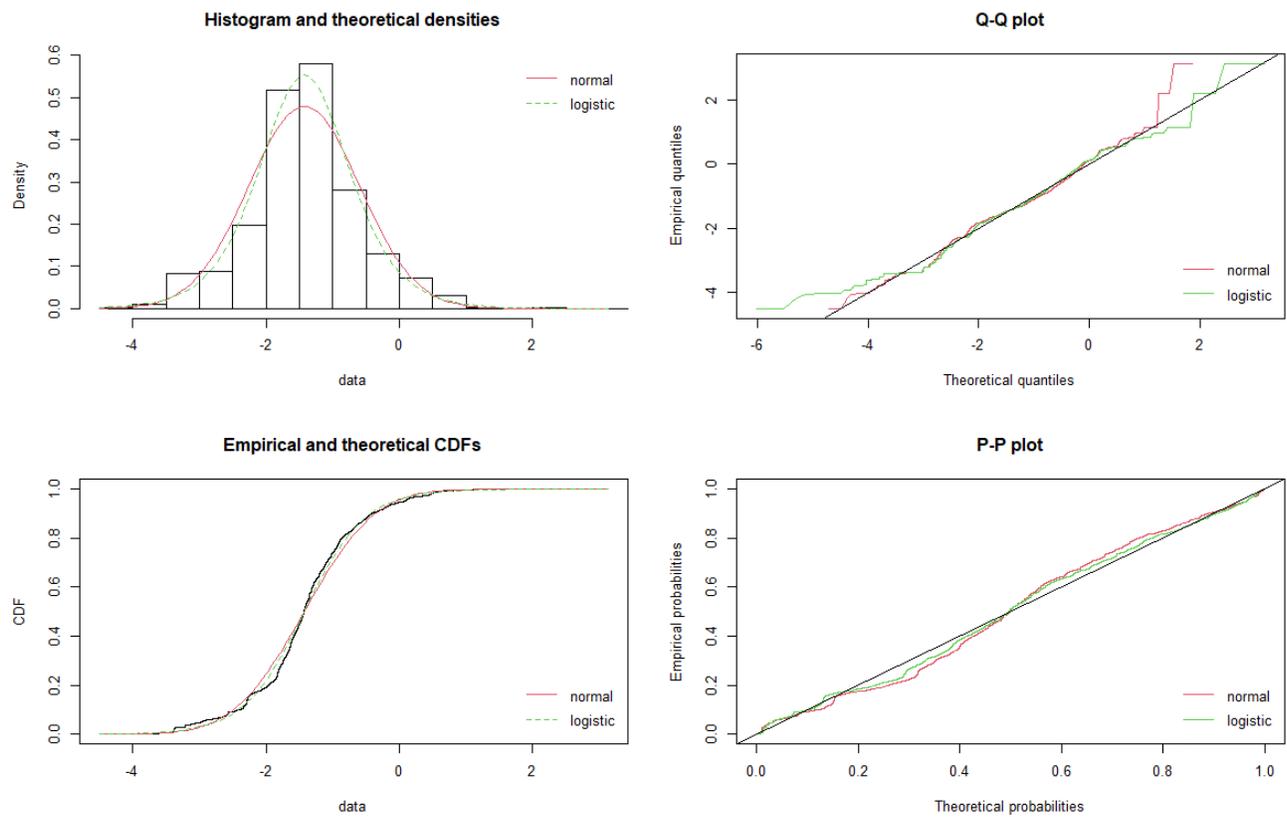


Figure 9: Integrated visual analysis of **standardized measure of rainfall (z-score)**: distribution comparison and probability plots. This figure shows that Logistic distribution provides a comparatively better fit for the data.

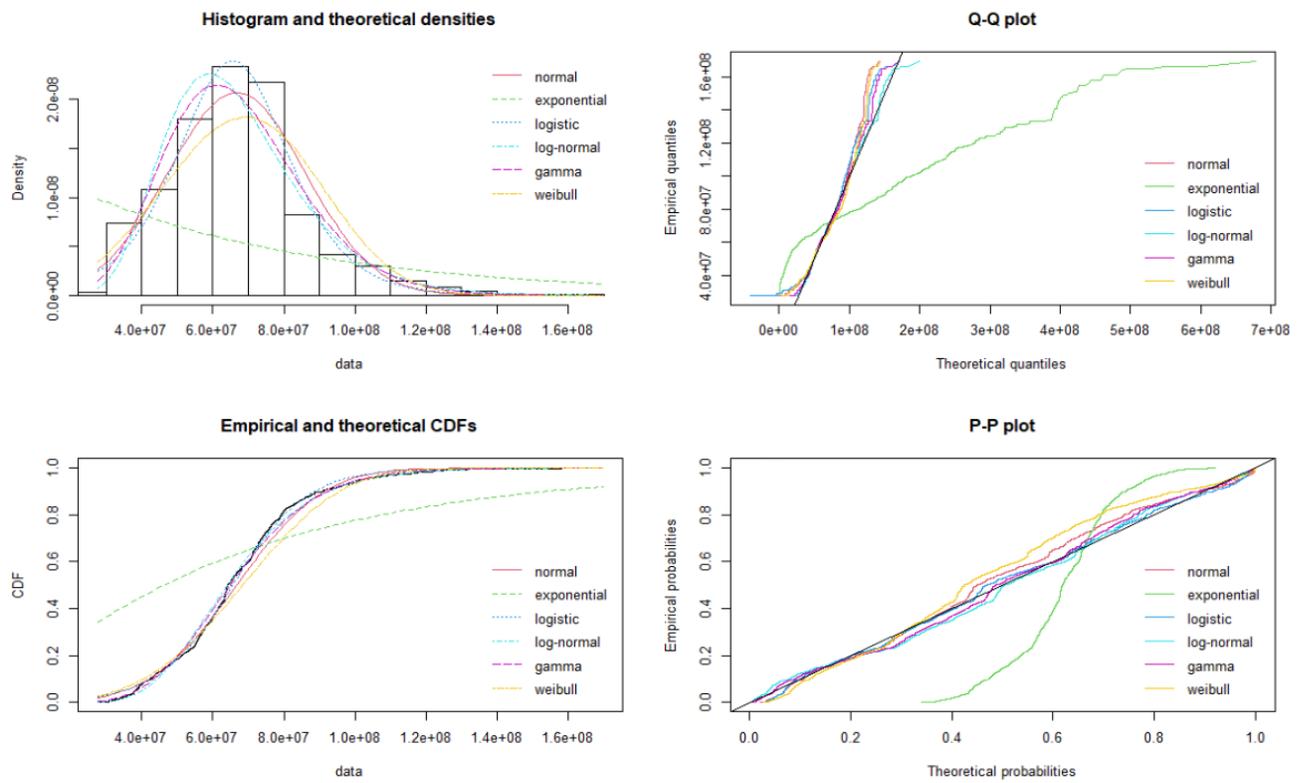


Figure 10: Integrated visual analysis of **vegetation index**: distribution comparison and probability plots. This figure shows that except for the Exponential distributions, almost all other selected ones provide a good fit for the data.

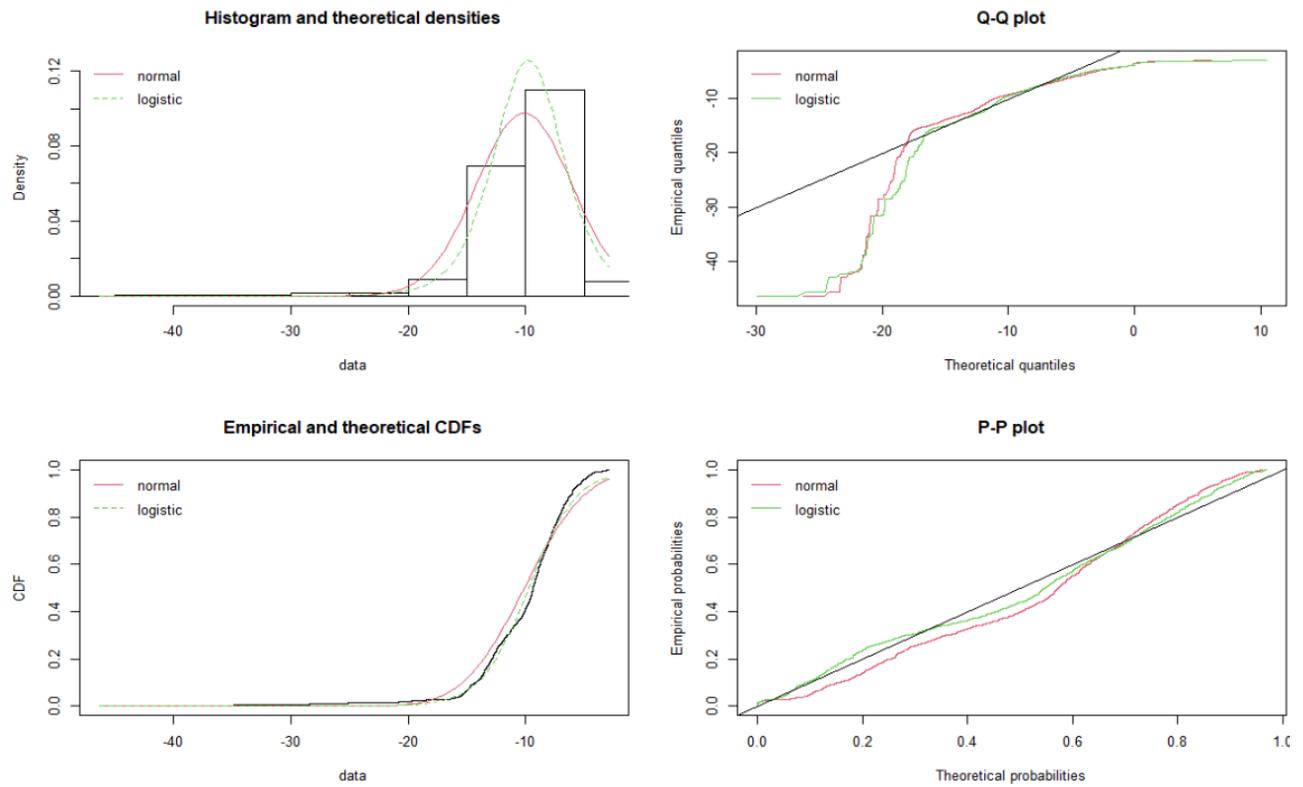


Figure 11: Integrated visual analysis of **standardized measure of vegetation index (z-score)**: distribution comparison and probability plots. This figure shows that Logistic and Normal distributions provide a similar fit for the data.

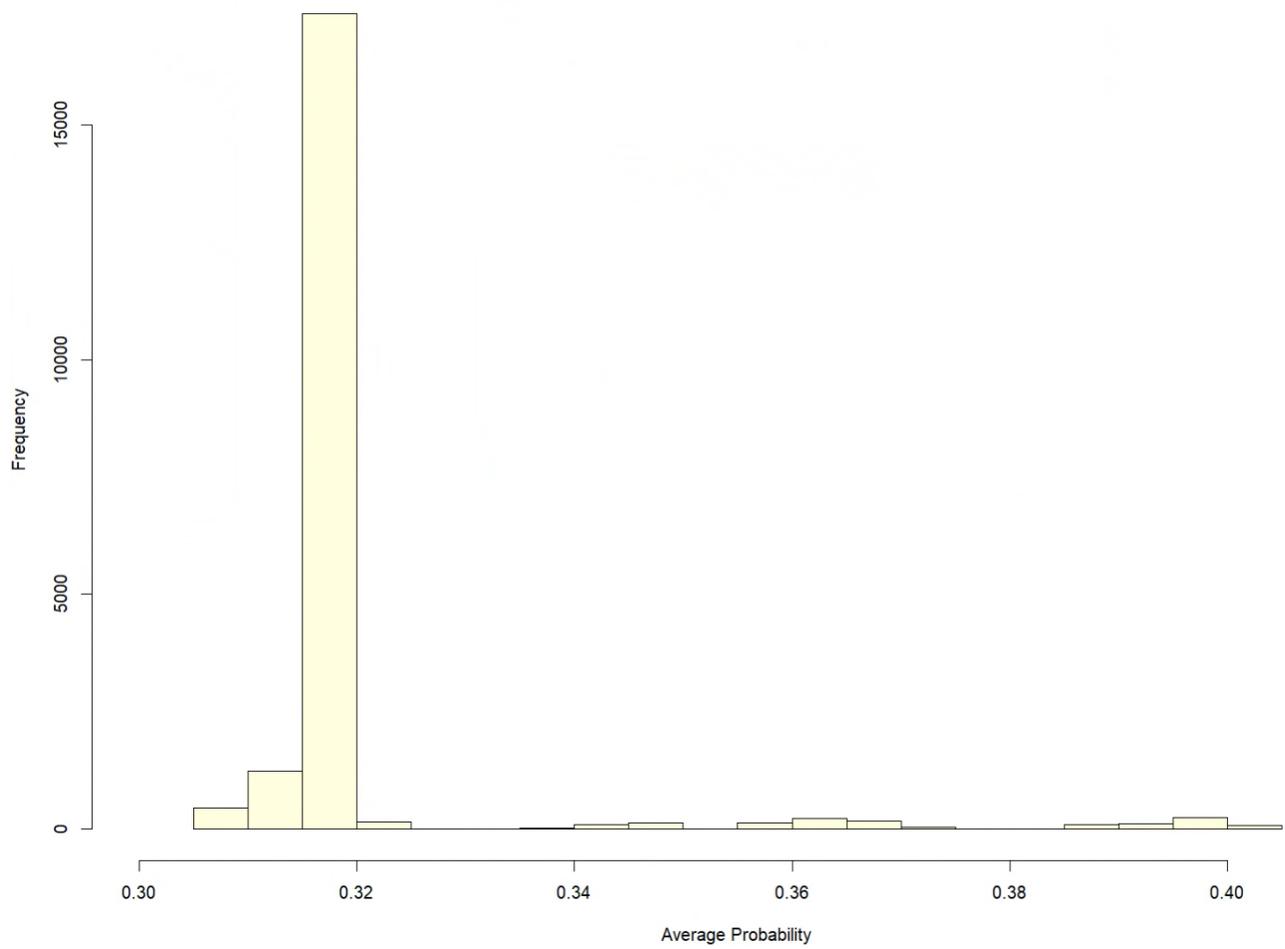


Figure 12: Summary of vulnerability results from ~20,500 estimated models. For the majority (73%) of the estimated models, the average vulnerability estimate is between 31 - 32%.

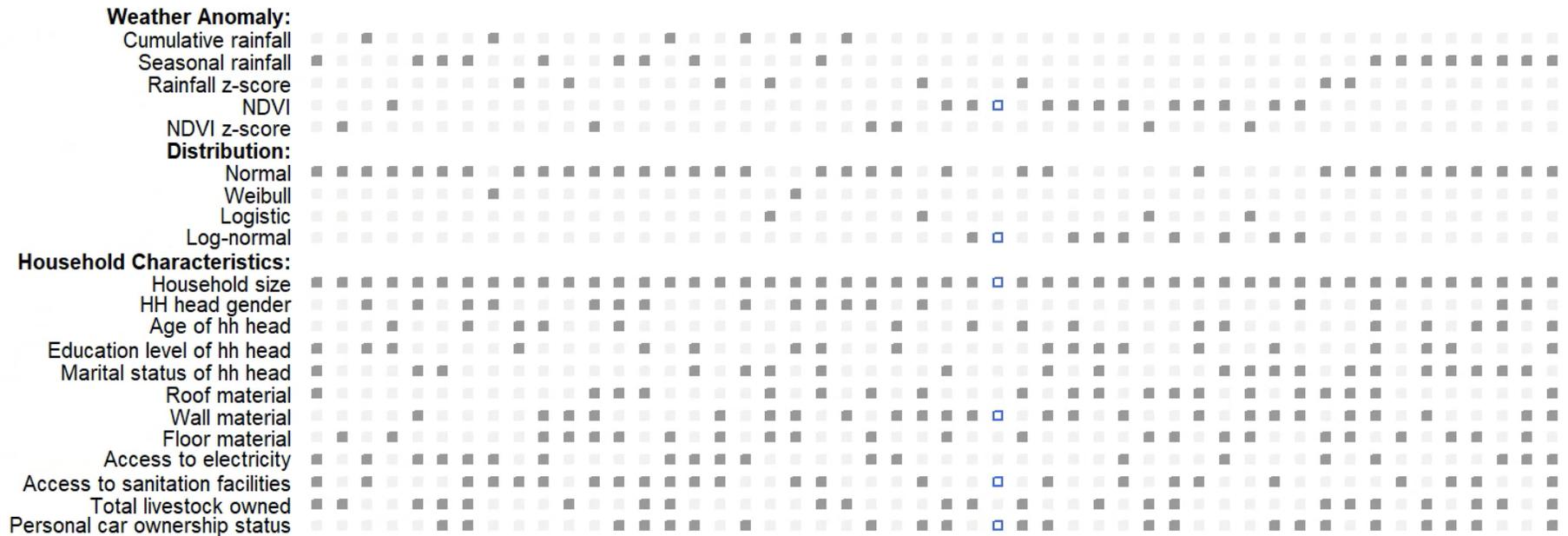


Figure 13: Results of vulnerability analysis. We present the results from randomly selected 50 models out of the pool of 20,500 estimated models.

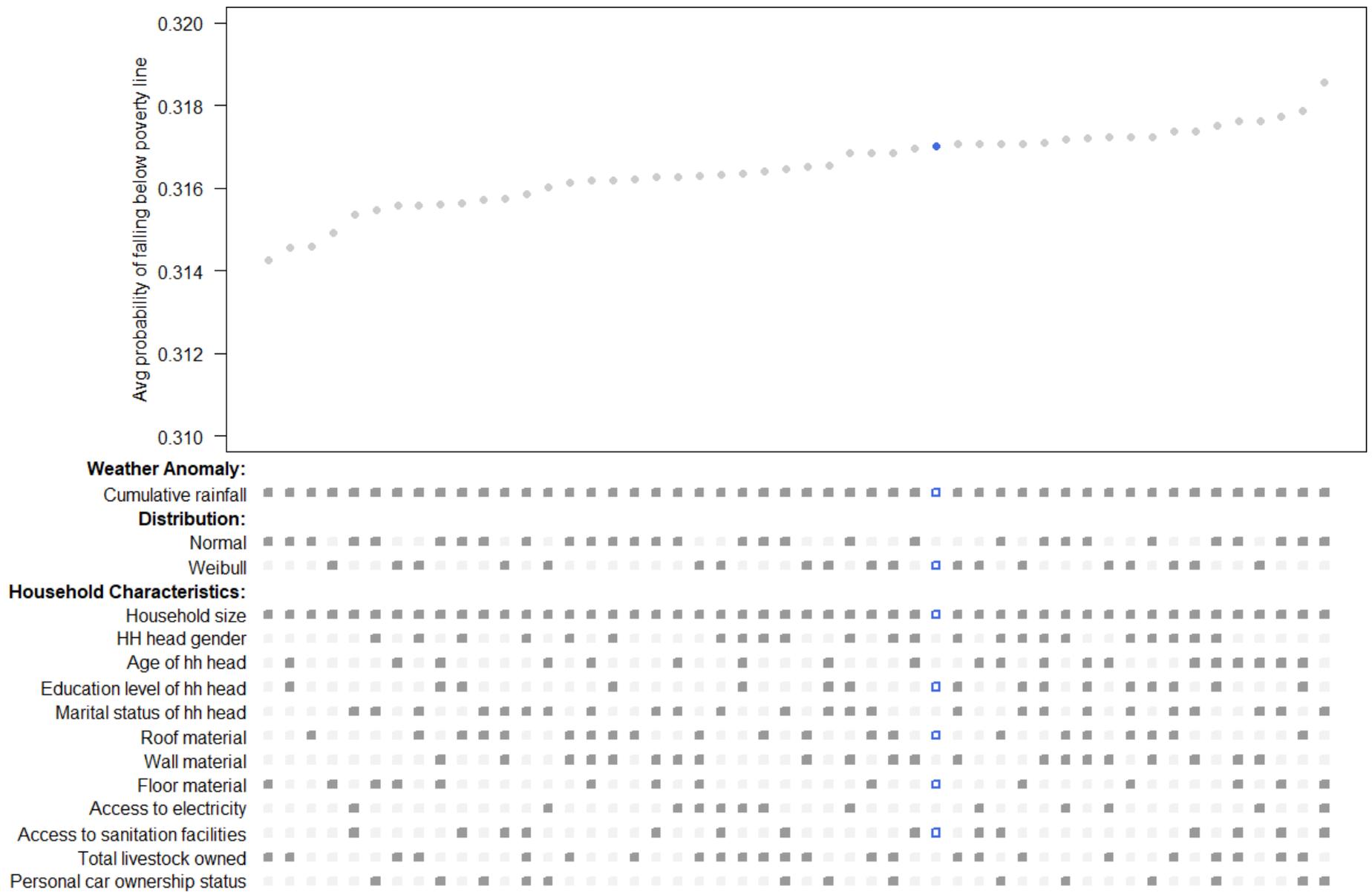


Figure 14: Results of vulnerability analysis for rainfall data.

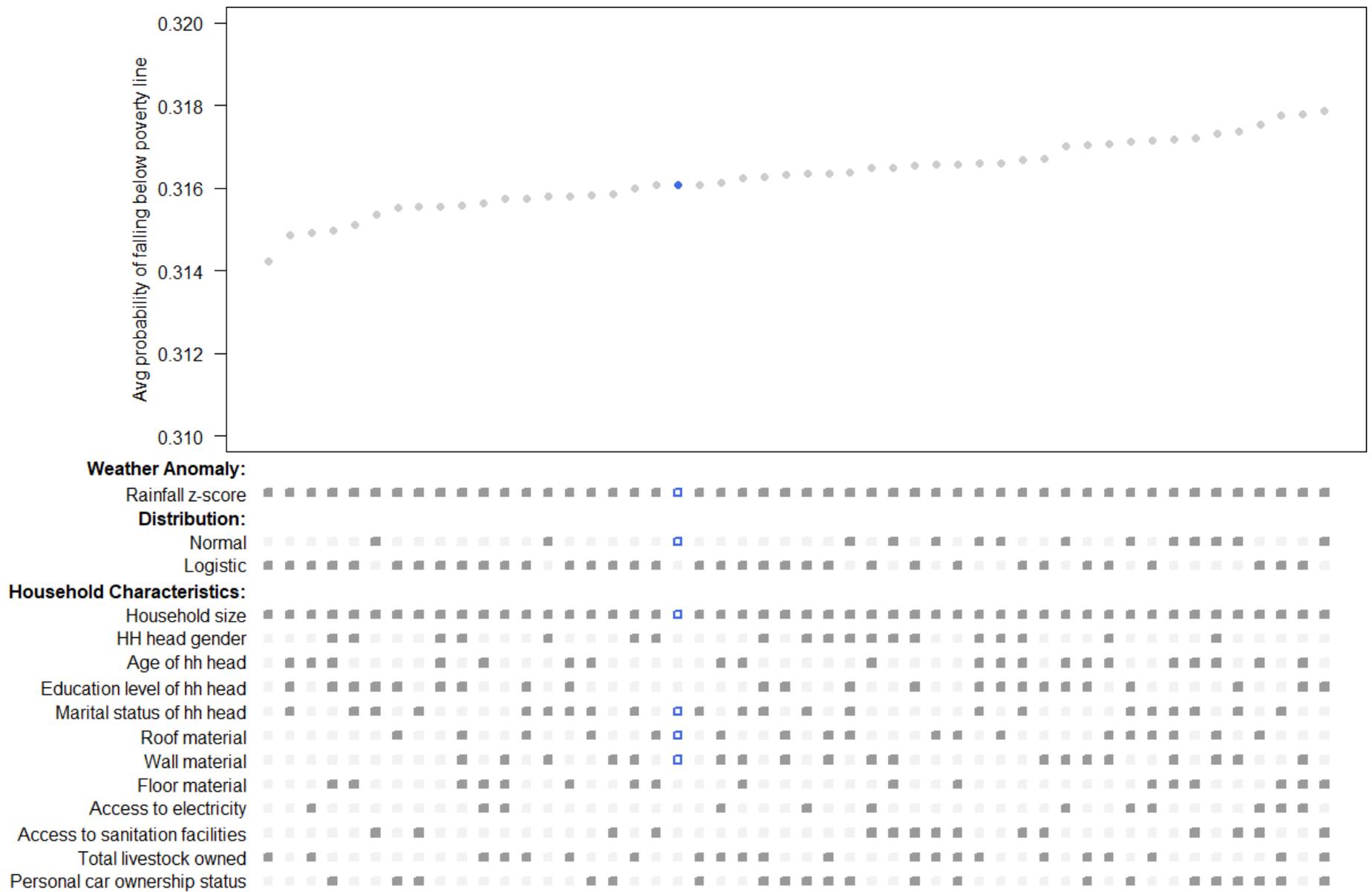


Figure 15: Results of vulnerability analysis for the standardized measure of rainfall.

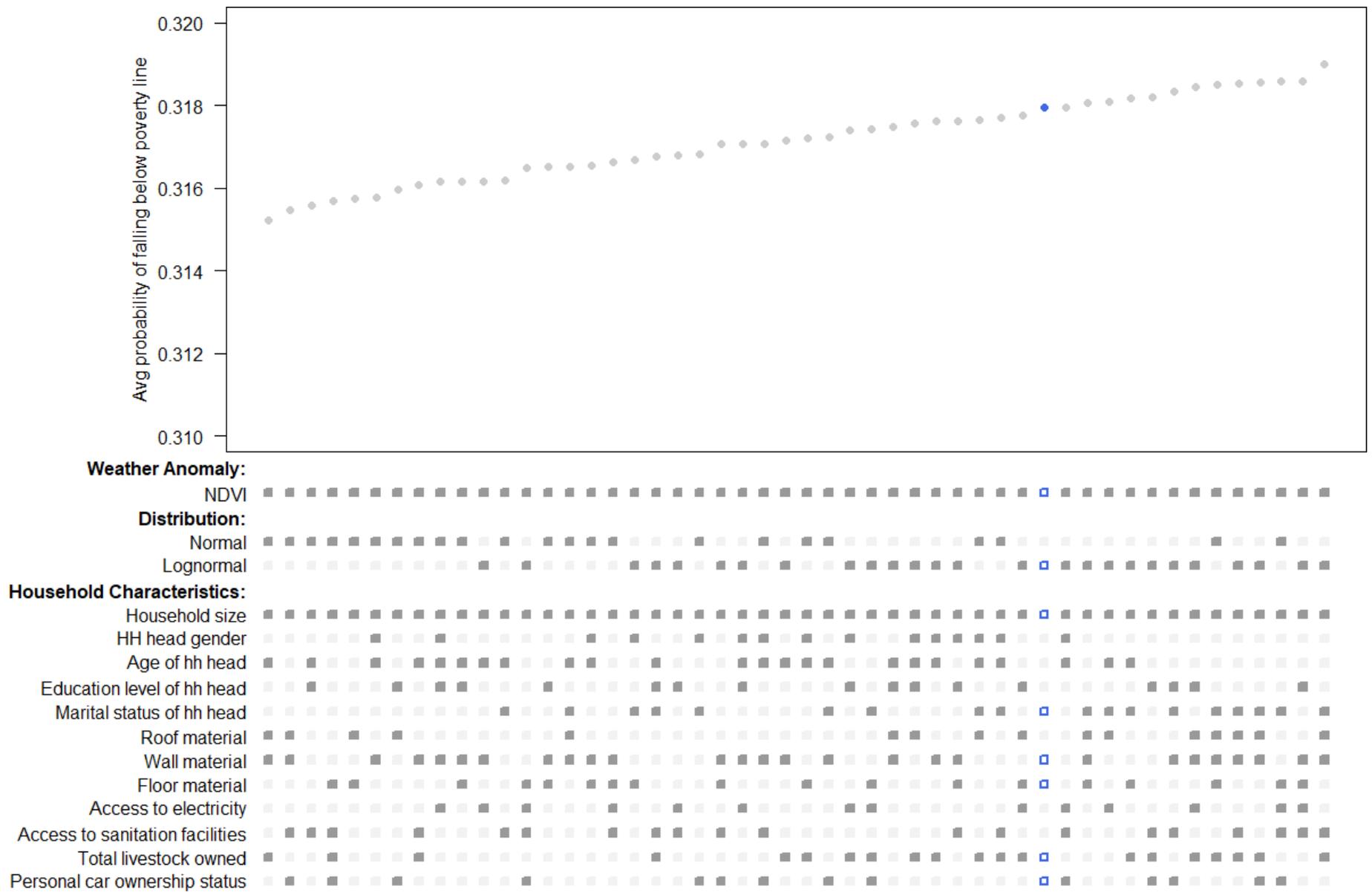


Figure 16: Results of vulnerability analysis for NDVI.

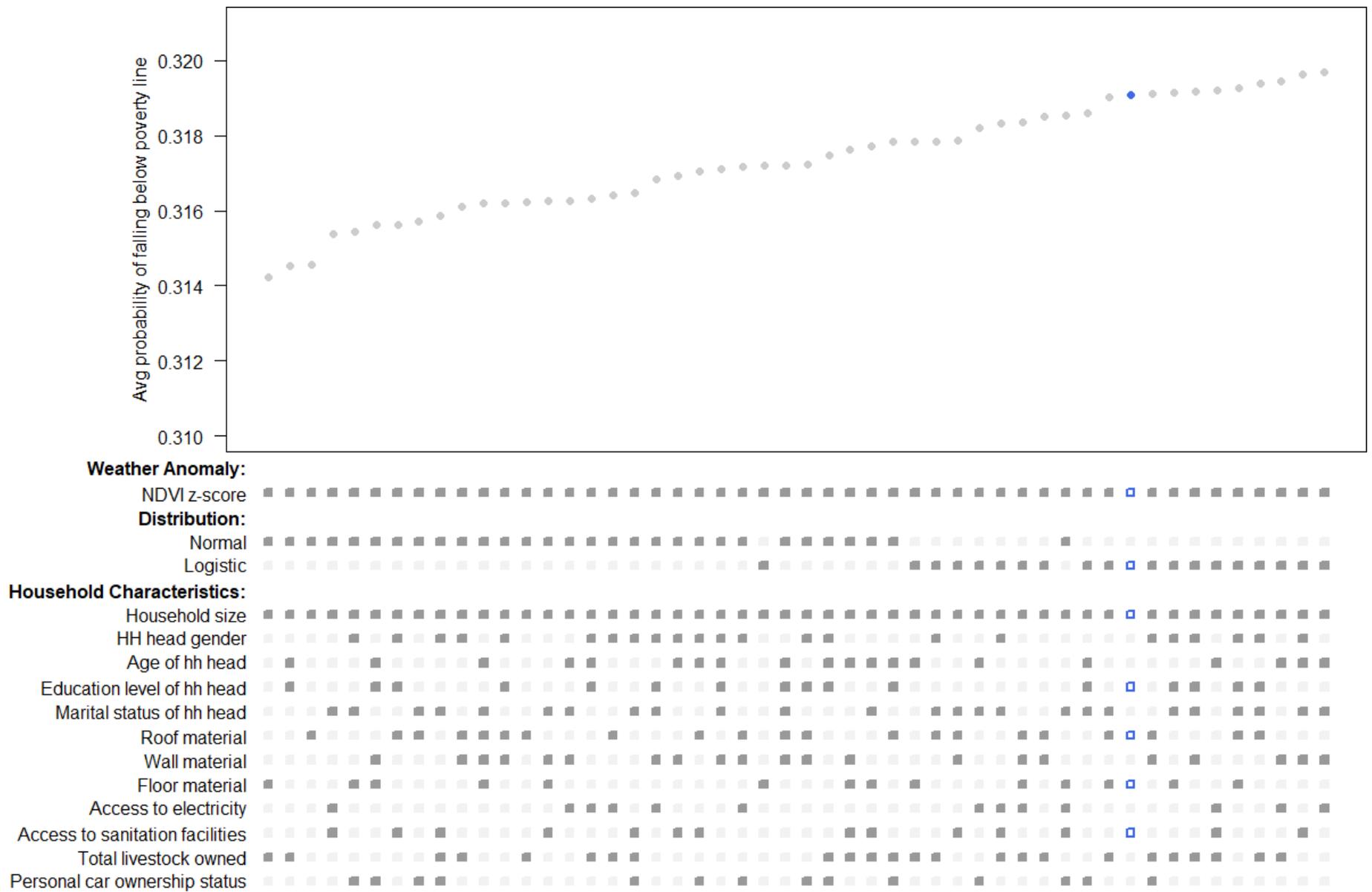


Figure 17: Results of vulnerability analysis for the standardized measures of NDVI.

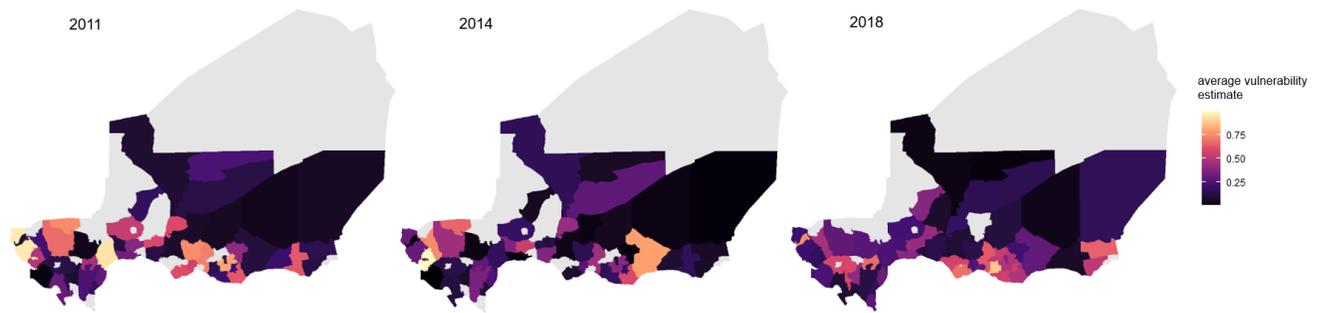


Figure 18: Average vulnerability estimate across space. The most vulnerable non-poor rural households reside in the south and south-western part of Niger.

Tables

Table 1: Descriptive Statistics of LSMS Data (in 2011 US\$, PPP)

Panel A. Summary statistics of consumption data for urban and rural departments					
Variable	Number of Obs.	Mean	St. Dev.	Min	Max
Consumption per capita in 2011 (\$)	3,519	1,383	1,185	180	18,135
Consumption per capita in 2014 (\$)	3,301	1,572	1,306	193	20,274
Consumption per capita in 2018 (\$)	6,012	1,426	1,558	160	41,519
Consumption per capita (full sample) (\$)	12,832	1,452	1,402	160	41,519

Panel B. Summary statistics of consumption data only for rural departments					
Variable	Number of Obs.	Mean	St. Dev.	Min	Max
Consumption per capita in 2011 (\$)	2,083	924	467	180	4,865
Consumption per capita in 2014 (\$)	2,067	1,113	646	193	9,754
Consumption per capita in 2018 (\$)	4,447	1,038	698	160	12,274
Consumption per capita (full sample) (\$)	8,597	1,028	640	160	12,274

Table 2: Descriptive Statistics for Agro-environmental Data

Variable	Number of Admin Unit	Mean	St. Dev.	Min	Max
Total precipitation in 2011 (mm)	50	316	124	19	535
Total precipitation in 2014 (mm)	50	353	154	20	656
Total precipitation in 2018 (mm)	61	441	183	21	940
Total precipitation (full sample) (mm)	63	375	166	19	940
Deviation from historical normal	63	0.36	0.86	-1.66	3.57
Peak NDVI in 2011	50	0.28	0.08	0.11	0.44
Peak NDVI in 2014	50	0.28	0.08	0.11	0.45
Peak NDVI in 2018	61	0.31	0.09	0.13	0.53
Peak NDVI (full sample)	63	0.29	0.08	0.11	0.53
Deviation from historical normal	63	0.04	0.76	-1.73	2.46

Note: The second column shows the number of unique administrative 2 units, i.e. departments, for which we have LSMS survey data.

Table 3: Definition and Summary Statistics of Household Characteristics

Covariate Name	Type	Description	Mean	Std
Household size	Continuous	What is the household size?	6.106	3.181
Age of the household head	Continuous	What is the age of the household head in years?	44.952	15.065
Gender of the household head	Binary	What is the sex of the household head?	0.845	0.362
		1 "male"		
		0 "female"	0.155	0.362
Education level of household head	Categorical	What is the highest education level attained by the household head?	0.870	0.337
		1 "no education"		
		2 "primary (complete or incomplete)"	0.103	0.304
		3 "secondary (complete or incomplete)"	0.024	0.153
		4 "tertiary complete or incomplete"	0.003	0.059
Marital status of household head	Categorical	What is the marital status of the household head?	0.684	0.465
		1 "married monogamous"		
		2 "married polygamous"	0.208	0.406
		3 "single/never married"	0.005	0.069
		4 "living together"	0.000	0.000
		5 "divorced or separated"	0.058	0.234
		6 "widowed"	0.045	0.207
Roof	Categorical	What is the main material used for the roof of the house?	0.243	0.429
		1 "cement/concrete/stones"		
		2 "tile/bricks"	0.069	0.253
		3 "metal-based materials"	0.210	0.407
		4 "thatch/straw"	0.162	0.368
		5 "earth-based materials"	0.314	0.464
		6 "other"	0.002	0.040
Floor	Categorical	What is the main material used for the floor of the house?	0.039	0.194
		1 "cement/concrete"		
		2 "tiles/wood"	0.001	0.031
		3 "earth-based materials"	0.952	0.214
		4 "other"	0.007	0.087
Wall	Categorical	What is the main material used for the external walls of the house?	0.012	0.108
		1 "cement/concrete/stones"		
		2 "bricks"	0.061	0.240
		3 "metal-based materials"	0.456	0.499
		4 "earth-based materials"	0.462	0.499
		5 "other"	0.009	0.093
Access to electricity	Binary	Does the household have access to electricity?	0.032	0.176
		1 "yes"		
		0 "no"	0.968	0.176
Access to sanitation facilities	Binary	Does the household have access to a toilet?	0.025	0.156
		1 "yes"		
		0 "no"	0.975	0.156
Total livestock owned	Continuous	How many livestock does the household own?	9.122	21.435
Ownership of personal car	Binary	Does the household have a personal car?	0.003	0.057
		1 "yes"		
		0 "no"	0.997	0.057

Table 4: Summary of the findings from the goodness of fit assessment

	Cullen and Frey Plot	Probability Distributions	KS, CvM, and AD Statistic
Cumulative Rainfall	Gamma	Exponential	
	Log-normal	Gamma	Weibull
	Weibull	Weibull	
Rainfall z-score	Logistic	Logistic	Logistic
	Log-normal		
NDVI	Logistic	Log-normal	Log-normal
		Gamma	
		Logistic	
NDVI z-score	Log-normal	Logistic	Logistic
	Gamma		

Note: The KS, CvM, and AD stand for Kolmogorov-Smirnov, Cramer-von Mises, and Anderson-Darling statistic, respectively.