Do Zambian Farmers Manage Climate Risks?*

Soyoka Okamura[†] Ken Miura[‡] Chieko Umetsu[§]

March 1, 2024

Abstract

This study investigates the agricultural production responses to climate risk among small-scale farmers in Zambia by combining grid-level historical rainfall data with nationally representative household survey data. After identifying the importance of January and February rainfall in maize production, we define these two months' rainfall variations over the past 59 years as the climate risk index. We then relate this risk index to household-level agricultural decisions regarding risk diversification and farm investment. Results indicate little evidence of crop or plot diversification strategies in response to weather risks. Conversely, farmers in high-climate-risk regions apply less fertilizer and consequently achieve lower maize yields than their counterparts in low-risk regions. The mediation analysis attributes underinvestment in fertilizers to 38.5% of the rainfall risk impact on maize productivity. Overall, Zambian farmers manage climate risk by reducing risky farm inputs at the expense of returns.

JEL Classification: O12, O13, Q12.

Keywords: self-insurance, agriculture, weather risk, mediation analysis, Zambia

*We thank Niraj Prakash Joshi, Johannes Sauer, Hitoshi Shinjo, and the seminar participants at the technical workshop of the 11th Biennal Rosenberg International Forum on Water Policy, the Association for Regional Agricultural and Forestry Economics international workshop, and J-Tree for their valuable comments and suggestions. We are grateful to the Zambia Statistics Agency (ZamStats) staff for their generosity in sharing data. The authors acknowledge the financial support from JSPS KAKENHI Nos. 20H00440 and 22K14957. Any errors are our own.

[†]University of Tokyo. Email: okamura-soyoka134@g.ecc.u-tokyo.ac.jp

[‡]Kyoto University. Corresponding author: Division of Natural Resource Economics, Graduate School of Agriculture, Kyoto University, Kitashirakawa Oiwake-cho, Sakyo-ku, Kyoto 606-8502, Japan; Email: miura.ken.3e@kyoto-u.ac.jp

[§]Kyoto University. Email: umetsu.chieko.5e@kyoto-u.ac.jp

1 Introduction

The impact of climate change has become increasingly conspicuous worldwide. Developing countries are vulnerable to climate change, and because small-scale farmers rely mostly on rainfed agriculture, they are particularly exposed to severe weather risks (Kurukulasuriya et al., 2006). Understanding climate risk management is critical to designing effective and appropriate adaptation policies.

According to the standard economic theory, risk-averse agents are willing to diversify their income risk in risky environments. Where the credit and insurance markets are underdeveloped, the most accessible risk diversification is the adjustment of income portfolios by increasing investments in low-risk assets in exchange for higher returns. Similarly, agents are likely to avoid profitable, albeit risky, investments. Although ensuring a secure income is a critical livelihood strategy for individuals living close to a subsistence level to bypass life-threatening scenarios in the short run, missing profitable opportunities may lock them into poverty traps in the long run. To derive welfare implications, this study investigates the nature of risk diversification as self-insurance and its consequences for productivity in Zambian agriculture.

Zambia provides an excellent setting for empirical analysis of farmers' risk management. First, agricultural production is prone to climate risks because irrigation facilities are almost nonexistent in rural areas; thus, farming is rainfed (Mendelsohn and Dinar, 2003). Second, the mono-production mode of maize crops continues to be dominant among Zambian smallholders despite the government's and aid organizations' efforts to promote crop diversification as a practical adaptation strategy against climate risks. Farmers' focus on maize production may be risky given the high weather risks, providing an empirical puzzle motivating this study. Therefore, investigating household risk management in agricultural production offers valuable implications for future policy planning.

In the literature, previous studies on self-insurance empirically examined the production response of agricultural households to climate risk. Examples of such agricultural decisions include crop choices, seed choices, land adjustments, and farm investments in fertilizers and labor (Alem et al., 2010; Karlan et al., 2014; Emerick et al., 2016; Arslan et al., 2018; Aragón et al., 2021; Boucher et al., 2021). However, few empirical attempts have been made to discuss the consequences of farmers' weather risk management practices on their agricultural productivity. To fill this research gap, this study examines how climate risk affects farmers' agricultural decisions and, consequently, farm productivity in rural Zambia by combining nationally representative agricultural survey data and long-term pixel-level climate data.

Our analysis begins by defining a climate risk index based on historical variations in rainfall amounts that are crucial for agricultural production. We estimate the impact of monthly precipitation on maize yield for each calendar month using district-level production records and rainfall estimates from 1990/91 to 2018/19 cropping seasons. Past production records were obtained from annual agricultural statistics aggregated at the district level using the Crop Forecast Survey (CFS) conducted by the Zambia Statistics Agency in collaboration with the Ministry of Agriculture. For historical rainfall data, we aggregate the grid-level rainfall database, WorldClim, at the district level. Using these data, our estimation results identify the rainfall in January and February as the most influential determinants of maize yield. Based on this result, we define the coefficient of variation in the rainfall of these two months over 59 agricultural years (1960/61 to 2018/19) as the climate risk index for this study and construct it at the ward level.

We then relate this climate risk index to agricultural decisions concerning risk diversification and farm investments of more than 10,000 farm households from nationally representative CFS data for the 2020/21 cropping season. As a suitable nature for this study, the CFS collected household-crop-plot level information on seed choices and fertilizer applications. This allows us to consider a wide range of risk management strategies and analyze them at extensive and intensive margins. The estimation results reveal no evidence that farmers diversify their planted crops or plot locations in response to climate risks. Additionally, we find little evidence of growing drought-tolerant crop varieties such as sorghum and millet in high-climate-risk regions. Conversely, the empirical results suggest that farmers respond significantly to climate risks by reducing fertilizer application and adopting hybrid maize seeds that are typically drought-tolerant and early-maturing and, thus, risk-hedging inputs, which is consistent with theoretical predictions. These results are also economically significant; a one-standard deviation increase in our rainfall risk measure reduces the fertilizer applied to the field by 14.9 kilograms per hectare, corresponding to approximately 13% of its standard deviation, and increases the likelihood of planting hybrid maize seeds by approximately ten percentage points, with a sample average of 74%. As our regressions control for recent weather shocks, these input responses directly capture the long-run behavioral reactions to location-specific rainfall risks.

Our findings that rainfall risk significantly changes household investment decisions in farming invite natural speculation that climate risk has consequences for productivity. The data indicate that, after accounting for soil conditions and recent climates, the maize yield gap is approximately 9% when the difference in our climate risk index equals one standard deviation. To quantify the cost of climate risks via household-level risk-management behavior, we conduct a mediation analysis to examine the extent to which the responses of fertilizer application and hybrid seed adoption to rainfall risks contribute to maize productivity (Acharya et al., 2016). Specifically, we estimate the average conditional direct effect of historical rainfall variations conditional on fertilizer and seed inputs and then compare these estimated coefficients to discuss the relative importance of these two channels. This empirical exercise demonstrates that risk-induced underinvestment in fertilizer reduces maize productivity by 38.5%, while encouraging hybrid seed adoption restores it by 65.8% in proportion to the total productivity loss owing to increased climate risks. Thus, risk avoidance through underinvestment in chemical fertilizers is costly for Zambian farmers, whereas risk hedging by planting hybrid seeds has positive productivity consequences as a by-product.

This study contributes to the literature on the impact of climate risk on the welfare of farmers in developing countries. Previous studies have examined land values, crop yields, and agricultural productivity (e.g., Kurukulasuriya et al., 2006; Welch et al., 2010; Lobell et al., 2011; Chen et al., 2016; Taraz, 2018) as welfare indicators influenced by climate change. Instead of estimating the reduced-form impacts of weather conditions, this study conducts a mediation analysis to uncover the impact of risk-induced household behavior as a channel through which climate risk affects agricultural productivity. Our closest study is that of Chen and Gong (2021). They use a county-year panel over the past 35 years in China and decompose the impact of climate change on crop yields into the effect of changes in total factor productivity and agricultural input utilization. Although both studies investigate the mechanisms underlying climate adaptation, they are distinct in two important ways. First, Chen and Gong (2021) unpack the impacts of climate change on agricultural outputs but do not quantify the relative importance of the two channels in productivity consequences. Our results indicate that the adoption of hybrid maize seeds has yield-enhancing effects. However, these favorable effects are attenuated by the negative effects of underinvestment in chemical fertilizers in response to rainfall risks. The second important difference is the unit of analysis. While Chen and Gong (2021) use aggregated data at the county level, this study examines plot-level data to obtain more precise estimates by controlling for important determinants of agricultural decisions such as household demographics and plot characteristics. This difference between the two studies may generate different findings on labor responses to climate factors, with no significant results in this study.[1]

Another contribution of this study is the addition of new evidence to the rich literature on smallholder household behavior in risky environments in developing countries. One strand of the literature identifies various agricultural decisions as a response to climate risks.[2] Among them, the study by Arslan et al. (2018) is notable. They examine the relationship between long-term precipitation risks and three types of diversification (crop, livestock, and income) in Zambia using different data sources from our study and find crop portfolio diversification as a response to rainfall risks in dry regions. In contrast to their findings, we find no significant risk management through diversification strategies among Zambian farmers. Although the varied results may be attributed to different data sources and empirical samples, the current assessment of farmers' risk management in Zambia requires further data collection and empirical investigation. Furthermore, our finding of no evidence for crop and plot diversification strategies warrants future investigations into the potential hindrances to traditional self-insurance in agricultural production.

The remainder of this paper is organized as follows: Section 2 describes the theoretical motivation for this study. Section 3 provides background information on Zambian agriculture and discusses the nature of the data used in the subsequent empirical analysis. After constructing the climate risk index used in this study in Section 4, Section 5 investigates household production responses to climate risks. Section 6 discusses the productivity consequences of risk-induced household production behaviors through a mediation analysis. Section 7 confirms the robustness of the primary results. Finally, Section 8 summarizes the findings and proposes a future research agenda.

[1] Both studies confirm household responses through fertilizer adjustment. In addition, our observed hybrid seed utilization and fewer fertilizer applications in high climate-risk areas may suggest that farmers adapt to uncertain environments by adopting agricultural technologies suitable for risky production, which is consistent with adaptations in the agricultural input portfolio, highlighted by the findings of Chen and Gong (2021).

[2] The examples include Dercon (1996) for crop choice, Aragón et al. (2021) and Liu et al. (2023) for land adjustment, Alem et al. (2010) for farm investments in fertilizers, Ito and Kurosaki (2009) for agricultural labor adjustment, and Cui and Xie (2022) for changing the growing season.

2 Risk Management in Agricultural Production

People in developing countries are vulnerable to unpredictable shocks owing to family illnesses and extreme weather conditions. Without any risk-coping strategies, these shocks often lead to detrimental effects on household welfare in general and irreversible consequences for human capital accumulation in particular. Given the salience of risks to their livelihoods, a natural hypothesis is that farmers form expectations regarding the probability distribution and magnitude of each shock based on their experience and local conditions. Additionally, their expectations were updated, if necessary, by observing the events in each period. While some risks are beyond farmers' control (e.g., rainfall risks), farmers can control the consequences in advance in two ways: by transacting risks with others and through self-insurance.

First, risk can spread across individuals; thus, uninsured risk creates a demand for risk pooling within groups. Informal risk pooling within social networks is pivotal in shielding household consumption from income fluctuations in rural economies.[3] A typical form of informal risk-sharing arrangement involves cash and gift transfers and labor exchanges with relatives and neighbors. Thus, a group that diversifies risk tends to be geographically confined, suggesting that informal arrangements can be more effective against idiosyncratic shocks (e.g., family illness) than aggregate shocks (e.g., climate shocks). Theoretically, to protect against aggregate risk, it is possible to spread it through formal markets. A leading example is commercialized insurance products for health and weather events. However, formal insurance markets in developing countries are immature, because the absence of market institutions that support smooth transactions makes contract enforcement extremely challenging. Full commitment to the initially agreed-upon contract is also difficult in informal risk-sharing arrangements, although peer pressure can help enforce the agreement. In addition to enforcement issues, classical information asymmetry problems such as moral hazard and adverse selection undermine the stability of arrangements and markets. Thus, some risks remain uninsured even after sharing them with others.

Second, the risk can spread across activities through self-insurance. As self-insurance methods

[3] The following three features of traditional village economies make informal insurance arrangements effective. First, close-knit relationships within a group make mutual monitoring effective and provide perfect information settings. Second, villagers have limited outside options and depend heavily on the community and its members for most of their lives, making social sanctions effective when contracts are breached. Third, long-term relationships with the group members create repeated game settings, making opportunistic behavior an irrational choice.

diversify the risks faced in household production, issues relating to asymmetric information and contract enforcement are less concerning than formal contract-based insurance and informal risk arrangements. Although its effectiveness remains an empirical question, self-insurance is the most accessible risk-hedging method for small-scale farmers in developing countries. Therefore, understanding the nature of self-insurance is indispensable for designing policies aimed at enhancing the resilience of people's livelihoods against unexpected shocks. Among the several forms of self-insurance,[4] this study examines *ex-ante* risk management in agricultural production. Specifically, we focus on diversification and investment choices.

Risk diversification through crop choice is a traditional risk-management strategy in agrarian settings. Agricultural production is inherently risky, primarily because of unforeseen climatic conditions. The risk to production becomes salient, particularly when agriculture is rainfed. With distinct production responses to weather conditions, each crop has different expected returns and variances (Kurukulasuriya et al., 2006). Thus, farmers select an optimal crop portfolio by balancing the tradeoff between expected profits and production risks, given their risk attitudes and the nature of the risks in their production environments. Crop diversification can reduce total production risk in the absence of a perfect yield correlation between crops (Newbery, 1991).

Similarly, plot diversification is another way to spread production risks within a production mode (Morduch, 1995). Farm households can minimize production loss from crop disease and livestock/bird attacks by planting the same crop on multiple plots. Although aggregate weather risks cannot be insured by nature, this risk management strategy is also effective if the microclimates are salient.

Changing the production mode to a safer one is an alternative risk-management strategy for agricultural production. Similar to crop types, returns on inputs respond differently to production risks. If the returns on investment in farm inputs respond negatively to shocks, risk-averse farmers hesitate to use these inputs. The leading example is fertilizer, because its net return is small when weather shocks (e.g., drought) occur. Thus, we hypothesize that fertilizer application decreases in areas with high climate risk.

In contrast, some inputs can contribute to hedging production risks. For instance, variance in

[4] Precautionary saving through the accumulation of liquid assets such as livestock and jewelry can provide effective self-insurance against negative income shocks in developing countries (Fafchamps et al., 1998; Miura et al., 2012). The data constraint did not allow us to investigate their empirical roles in this study.

the profits from planting drought-tolerant crops and seed varieties is lower than those from planting regular crops and varieties. Another example of reducing the variance is planting early-maturing varieties because quicker crop cycles can minimize the ill effects of erratic rainfall patterns and drought.[5] Overall, we add the positive response of planting drought-tolerant crops and early maturing seeds to climate risks to our empirical hypotheses.

Finally, the responses of land and labor investments to production risks are theoretically ambiguous. As land rental and labor costs are minimal where outside options are limited, the responses of investment returns may be neutral to climate shocks. Thus, the direction in which they respond to climate risks depends on their production relationships with other inputs, such as fertilizer. For example, if labor and fertilizer are complements (substitutes), weather risks discourage (encourage) farmers from applying labor. Therefore, the relationship between climate risks and investments in land and labor is an empirical question.

In summary, the theory suggests that farmers in high-climate-risk regions are more likely to (1) diversify crops, (2) diversify plots, (3) plant drought-tolerant crops and varieties, and (4) plant early-maturing varieties compared with their counterparts in low-risk regions. Moreover, farmers in high-climate-risk regions are less likely to apply chemical fertilizers than their counterparts, while predictions regarding labor and land inputs are ambiguous ex-ante. This study tests these empirical hypotheses by combining historical climate estimates with household production data from Zambia.

3 Context and Data

3.1 Context

As in the rest of the sub-Saharan African countries, Zambia is an agriculture-based country. More than half of the population lives in rural areas (54% in 2022), and agricultural employment accounted for 59% of the total employment in 2021. However, the value added from the agriculture, forestry, and fishing sectors accounted for 3.4% in 2022, suggesting that most farmers engage in

[5]Planting early maturing varieties would also work as a risk-coping mechanism in the aftermath of climate shocks because farmers can replant it after the first planting is unsuccessful, owing to dry spells in the early production stage.

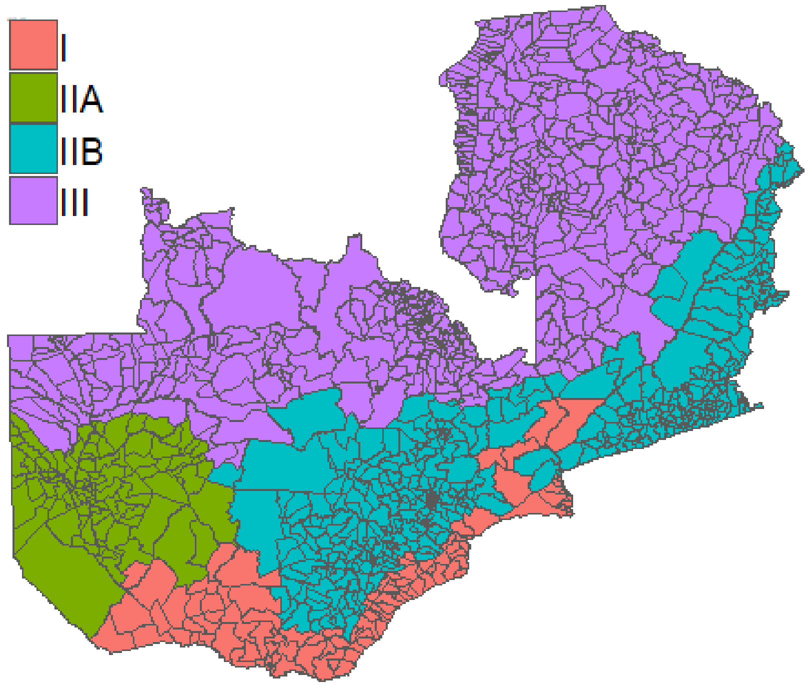

Figure 1: Agroecological zones in Zambia

Source: Shapefiles depicting this map are available at http://landscapesportal.org/layers/geonode:agroecological_zones for agroecological zones and at https://maps.princeton.edu/catalog/stanford-yc436vm9005 for ward boundaries.

subsistence agriculture.[6]

Environmental conditions are heterogeneous across the country. Based on the rainfall distribution and soil quality, the country is divided into three agroecological zones: I, II, and III (Ministry of Agriculture and Ministry of Fisheries and Livestock, 2016) (Figure 1). Region I in southern, eastern, and western Zambia, accounting for 12% of the country's total area, receives less than 800 mm of rainfall on average per year and has loamy to clayey soil on the valley floor and course and shallow loamy soils on the escarpment. Therefore, Region I is the driest zone, with frequent droughts. Region II accounts for 42% of the country, where the expected annual rainfall ranges

[6] All the statistics in this paragraph come from the World Bank Indicators (World Bank, 2022b, 2021, 2022a).

between 800 and 1000 mm, and is further divided into Region IIa with relatively fertile soils and Region IIb with sandy soils in the Western Province. Region III accounts for 46% of the country, and its average annual rainfall ranges between 1000 and 1500 mm. Despite the high rainfall, agricultural productivity is low because Region III has acidic soils caused by leaching.

Smallholders rarely have access to irrigation facilities in rural Zambia. The Food and Agriculture Organization (FAO) has estimated the proportion of land irrigated to total arable land as constant at 4–6% in the last two decades (2002–2019) (FAO AQUASTAT). Therefore, most agricultural production systems are rain-fed. As formal insurance and social safety nets are underdeveloped, weather shocks often depress food production and threaten national food security. Climate risk poses a major threat to Zambian farmers, and the government promotes investments in irrigation and crop diversification to enhance their resilience to climate change (ZVAC, 2015).

Although detailed crop calendars should be specific for each region, the main agricultural season, by and large, corresponds to the rainy season from November to April. Most farmers cultivate maize, which is a staple food in Zambia, during the rainy season. From the CFS in the 2020/21 agricultural season,[7] 94% of farmers cultivate maize, and approximately 80% of farmers cultivate only maize as cereal.[8] Despite salient weather risks, the dominant mono-production mode of maize characterizes the agriculture of rural farmers in Zambia, motivating this study. Conversely, during the dry season between May and October, agricultural activities are limited to winter maize and vegetable production in the wetlands, locally known as dambos, and riverbanks because it rarely rains.

Before proceeding, touching on other essential crops in risk management strategies within agricultural production is worthwhile. Specifically, millet and sorghum are more drought-tolerant than maize, because their water requirements are generally mild. Appendix Figure A.1 illustrates the relationship between total precipitation during the rainy season and crop yield at the district level by crop. As depicted in Appendix Figure A.1, yield responses to rainfall are milder for millet and sorghum than for maize. We expect a high prevalence of millet and sorghum planting in high-climate-risk regions.

[7]The detailed data description of the CFS will be provided shortly.

[8]Cereal includes maize, sorghum, rice, and millet.

3.2 Data

This study primarily uses historical rainfall and agricultural survey data. This subsection describes the data sources used in the study.

3.2.1 Historical Rainfall Data

The grid cell level monthly precipitation data used for this study is obtained from WorldClim for 1960–2018 (Fick and Hijmans, 2017; Harris et al., 2020).[9] The spatial resolution is 2.5 minutes ($\approx$ 21 km^2 at the equator). This study also uses daily rainfall estimates from TAMSAT v3.1 for 2019–2021 to make up rainfall records for the recent period (Maidment et al., 2014; Tarnavsky et al., 2014; Maidment et al., 2017).[10] The spatial resolution of the TAMSAT data is 0.0375 degrees ($\approx$ 16 km^2).

By mapping these historical rainfall data onto administrative boundary data in Zambia, we calculate monthly precipitation at the ward level as its weighted average using the area of pixels as the weight.[11] Based on constructed ward-level monthly rainfall estimates over 59 years between the 1960/61 and 2018/19 cropping seasons, we define the rainfall risk index for each ward. The empirical analysis also uses average annual precipitation over 59 years and recent rainfall estimates at the ward level as control variables in the regression.

3.2.2 Agricultural Survey Data

The primary empirical analysis uses household data from the CFS for the 2020/21 agricultural year. The CFS is conducted by the Zambia Statistics Agency (ZamStats) in collaboration with the Ministry of Agriculture during March and April every year to provide a basis for inferring national food security in a given agricultural season. The CFS covers all provinces and provides a nationally representative sample through a two-stage stratified cluster sample design to select

[9]The historical monthly weather data from the WorldClim database is the CRU-TS 4.06 (Harris et al., 2020) downscaled with WorldClim 2.1 (Fick and Hijmans, 2017). These data are publicly available at https://www.worldclim.org/data/monthlywth.html.

[10]We use the original "rfe" variable without recovering and filling in the missing data.

[11]The country has ten provinces: Central, Copperbelt, Eastern, Luapula, Lusaka, Muchinga, Northern, North-Western, Southern, and Western. As administrative units, each province is divided into districts, which are further subdivided into constituencies and wards.

interviewed households. First, a sample of the Enumeration Areas (EA) is selected in proportion to the number of households based on the 2010 Census of Population and Housing.[12] In practice, the sampling procedure selected 680 Census EAs across the country for the 2020/21 CFS.

Stratification was based on the total crop area. The CFS classifies households cultivating less than five hectares as small-scale farmers, households cultivating between 5 and 20 hectares as medium-scale farmers, and households cultivating more than 20 hectares as large-scale farmers. In the second stage of the sampling, after listing all the households for each selected EA, 20 households cultivating less than 20 hectares (i.e., "small- and medium-scale" farmers) are randomly sampled for survey interview from each list.[13] Thus, the CFS targets 13,600 households every agricultural year. The 2020/21 CFS interviewed 13,553 households.

The CFS questionnaire starts with each household member's basic demographic characteristics. The agricultural module then collects detailed information on plot characteristics, farming practices (e.g., tillage methods), inputs such as seeds and fertilizers, and the expected production and sales for each field and crop in the corresponding agricultural season. As CFS interviews are usually conducted before the harvest is fully completed, the harvested quantities are recorded as self-reported and estimated values by the respondents.

To investigate farmers' risk-management behavior, we calculate outcome variables at the household level based on plot-level data. The outcomes of interest include crop-specific yields in quantity per hectare; risk diversification indices, such as the number of crops cultivated and the Gini-Simpson index of crop-specific areas; and per-hectare quantities of farm inputs, such as chemical fertilizer and labor. To aggregate information on seeds at the household level, we calculate the weighted average of hybrid seed indicators across maize plots, using the area of the planted field as the weight. Finally, the ward-level historical precipitation records from WorldClim are assigned to each CFS household.[14]

[12]The EA is the geographical unit used by ZamStats. At the time of the 2010 Population and Housing Census, ZamStats demarcated each ward such that each EA had 60–120 (80–150) households in rural (urban) areas. The sampling frame for the Census contained 25,631 EAs. These EAs were used as sampling frames for the CFS after 2010.

[13]Conversely, "large-scale" farmers are always captured by the CFS every year. While the CFS interviewed approximately 1600 large-scale farms in a separate survey, this study excludes this category from the analysis and focuses on smallholders' risk-management behavior.

[14]We use ward boundary data as of 2010 because the 2020/21 CFS relies on EAs from the 2010 census as the

Before conducting the main empirical analysis relating climate risks to household production behavior, we define a climate risk index for each region by estimating the district-level relationship between rainfall and agricultural production. For this purpose, we use district-level production data aggregated from household-level CFS data. In particular, the expected quantity of harvest and planted area aggregated at the district level have been published since 1990. Using district-level data between the 1990/91 and 2018/19 cropping seasons, Section 4 estimates the maize yield as a function of monthly rainfall at the district level.

3.3 Summary Statistics

Table 1 presents the descriptive statistics of the outcomes and the main explanatory variables used in the empirical analysis. These variables are at the household level. Table 1 indicates that the average household cultivates approximately three crops in more than three fields. Only 9% and 6% of the sample households grew millet and sorghum, respectively, during the 2020/21 rainy season. While the average amount of fertilizer is 50 kg/ha for both types, high standard deviations suggest significant variations in fertilizer application in the sample. Finally, three to four adult family members work on farming in an average household. We use these variables as outcomes and relate them to the rainfall risk index defined in the next section.

4 Constructing Rainfall Risk Index

This section defines the index used to quantify rainfall risk. To this end, we first examine which monthly rainfall significantly impacts maize yield, using historical rainfall and production data at the district level. We then calculate the long-term variability in monthly rainfall that is important for farming and define this as the rainfall risk index.

---

sampling unit, and geographical information, such as wards and constituencies, refers to information from the 2010 census. We could not match the CFS data from a few wards with the historical rainfall data because of a mismatch between the 2010 ward boundary data and the information provided in the CFS. The main analysis omits observations that could not be linked to the precipitation data.

Table 1: Summary statistics

Variable	Mean	Std. Dev.	Min	Max	Obs
CoV (Jan, Feb)	0.19	0.07	0.09	0.32	12220
CoV (Nov–Apr)	0.16	0.04	0.09	0.24	12220
Number of crops	2.90	1.60	1.00	16.00	12220
Gini simpson index	0.42	0.26	0.00	0.92	12220
Number of plot	3.20	1.70	1.00	16.00	12220
Cultivate millet = 1	0.09	0.29	0.00	1.00	12220
Cultivate sorghum = 1	0.06	0.23	0.00	1.00	12220
Basal fertilizer (kg/ha)	50.00	59.00	0.00	1675.00	12220
Top dress fertilizer (kg/ha)	50.00	61.00	0.00	1750.00	12220
Total fertilizer (kg/ha)	100.00	118.00	0.00	3425.00	12220
Hybrid maize seed	0.74	0.43	0.00	1.00	12220
Area planted/Area field	0.91	0.19	0.02	1.00	12220
Number of family labor	3.30	1.80	0.00	14.00	12220
Number of female family labor	1.60	1.10	0.00	10.00	12220
Number of male family labor	1.70	1.20	0.00	9.00	12220

Notes: CoV of Prec (Jan, Feb) is calculated at the ward level and assigned to each household. Fertilizer and hybrid maize seed variables were originally at the plot level and were converted to household-level variables by calculating weighted averages with plot areas as the weight. The other variables are originally at the household level.

4.1 Specifications

We specify the district-level relationship between the maize yield and monthly rainfall as follows[15]:

$$Yield_{dt} = \beta_1 R_{dt}^{Jan} + \beta_2 R_{dt}^{Feb} + \cdots + \beta_{11} R_{dt}^{Nov} + \beta_{12} R_{dt}^{Dec} + \beta_l t + \beta_q t^2 + \delta_p + \epsilon_{dt} \tag{1}$$

where $Yield_{dt}$ represents maize yield (defined as expected maize harvest quantity divided by area planted) in tons per hectare of district d in agricultural year t and R_{dt}^m is the rainfall amount of district d in month m of agricultural year t. Thus, regression model (1) assumes that all monthly rainfall amounts affect maize yields additively and linearly. We include linear and quadratic time trends t and t^2 to control for agricultural technological progress over the study period. δ_p represents province fixed effects which capture time-invariant geographic features and ϵ_{dt} is an error term.

In general, we can define November and December as the planting season and January and February as the weeding season, based on the crop calendar in Zambia. Using season-specific rainfall variables, we also specify and run the following regression equation:

$$Yield_{dt} = \beta_P R_{dt}^{Plant} + \beta_W R_{dt}^{Weed} + \beta_{PW} R_{dt}^{Plant} \times R_{dt}^{Weed} + \beta_3 R_{dt}^{Mar} + \beta_4 R_{dt}^{Apr} + \beta_l t + \beta_q t^2 + \delta_p + \epsilon_{dt} \tag{2}$$

where R_{dt}^{Plant} (R_{dt}^{Weed}) is rainfall during the planting (weeding) season in district d during agricultural year t. In addition to the independent effects on the maize yield in each season, we allow for the complementarity of rainfall across seasons by including their interaction term. Finally, we include March and April rainfall R_{dt}^{Mar} and R_{dt}^{Apr} as controls.

We run regression equations (1) and (2) using unbalanced panel data from 76 districts for 27 cropping seasons between 1990/91 and 2018/19.[16]

4.2 Results

Table 2 presents the estimation results for the regression equation (1) in Column (1) and equation (2) in Columns (2) and (3), respectively.[17] Regression results in Column (1) display positive and statistically significant impacts of rainfall in December, January, and February on maize yields.

[15]The purpose of this analysis is to assess the impact of rainfall rather than to estimate the specific form of the production function.

[16]District-level production data for 1998/99 and 2015/16 are unavailable. Thus, we use data for 27 agricultural years.

[17]Appendix Table A1 presents summary statistics for empirical variables used for the estimation.

Table 2: Rainfall and maize yield, 1990/91 and 2018/19

	All	Season	Season interaction
Prec Nov	0.56		
	(0.56)		
Prec Dec	0.85*		
	(0.48)		
Prec Jan	0.71**		
	(0.35)		
Prec Feb	3.22***		
	(0.37)		
Prec Plant		-0.091	1.08***
		(0.22)	(0.34)
Prec Weed		1.11***	2.18***
		(0.19)	(0.27)
Prec Plant × Prec Weed			-1.95***
			(0.50)
Prec Mar	-0.21	0.69	0.45
	(0.42)	(0.44)	(0.44)
Prec Apr	-1.61**	-1.50**	-1.43**
	(0.63)	(0.66)	(0.67)
Linear trend in year	-0.061***	-0.050***	-0.050***
	(0.012)	(0.012)	(0.012)
Square trend in year	0.0029***	0.0028***	0.0027***
	(0.00042)	(0.00041)	(0.00042)
Adj. R-Squared	0.39	0.38	0.39
Observations	1734	1734	1734

Notes: Robust standard errors clustered by district are reported in parentheses. Province fixed effects are included, but not reported. We control for the precipitation in May, June, July, August, September, and October in Column 1. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

By comparing the magnitudes of the estimated coefficients, the results also suggest that February rainfall has the most significant effect on maize production.

The estimation results in Column (2), which do not include the interaction term of seasons, suggest the relative importance of weeding season rainfall (January-February) compared with planting season rainfall (November-December). The amount of rainfall during the weeding season presents positive and significant correlations with maize yields, whereas the planting season rainfall indicates a null association after controlling for weeding season rainfall. Adding the interaction terms of rainfall from the two seasons confirms the substitutability of rainfall impacts across seasons (Column 3). Although this interaction effect is interesting, a more important observation is that the independent impact of weeding season rainfall is approximately twice as significant as that of planting season rainfall. Appendix Tables A2 and A3 confirm that neither adding year dummies instead of time trends nor including fixed effects for districts instead of provinces qualitatively change the results.

Overall, estimating the maize production function using historical data suggests the crucial role of weeding season rainfall in maize yield. This finding is consistent with field observations and previous studies (Waldman et al., 2017). Even if drought hits in the early stage of the rainy season, local farm households can replant early maturing seed varieties to offset losses. Conversely, erratic dry spells during the weeding season significantly limit crop growth, leading to poor maize harvest. Based on the observed relative importance of weeding season rainfall to planting season rainfall, the coefficient of variation (CV) for January and February rainfall is defined as the precipitation risk index.[18]

Figure 2 plots the coefficient of variation of January and February rainfall for 59 years between the 1960/61 and 2018/19 cropping seasons (left), the average Gini-Simpson index of areas planted by crop as the crop diversification indicator (right), and the average amount of fertilizer applied in kilograms per hectare (middle) for each constituency. The left panel presents high rainfall variations in the southern part of the country corresponding to Region I of the agroecological zone classification. While the relationship between the crop diversification indicator and rainfall

[18]Throughout all the specifications in Table 2, April rainfall is negatively correlated with maize yields. This consistent finding raises the risk of focusing only on weeding season rainfall. To ease this concern, we also use the historical variation of annual rainfall amounts between November and next April as an alternative rainfall risk index to check the robustness of the main results.

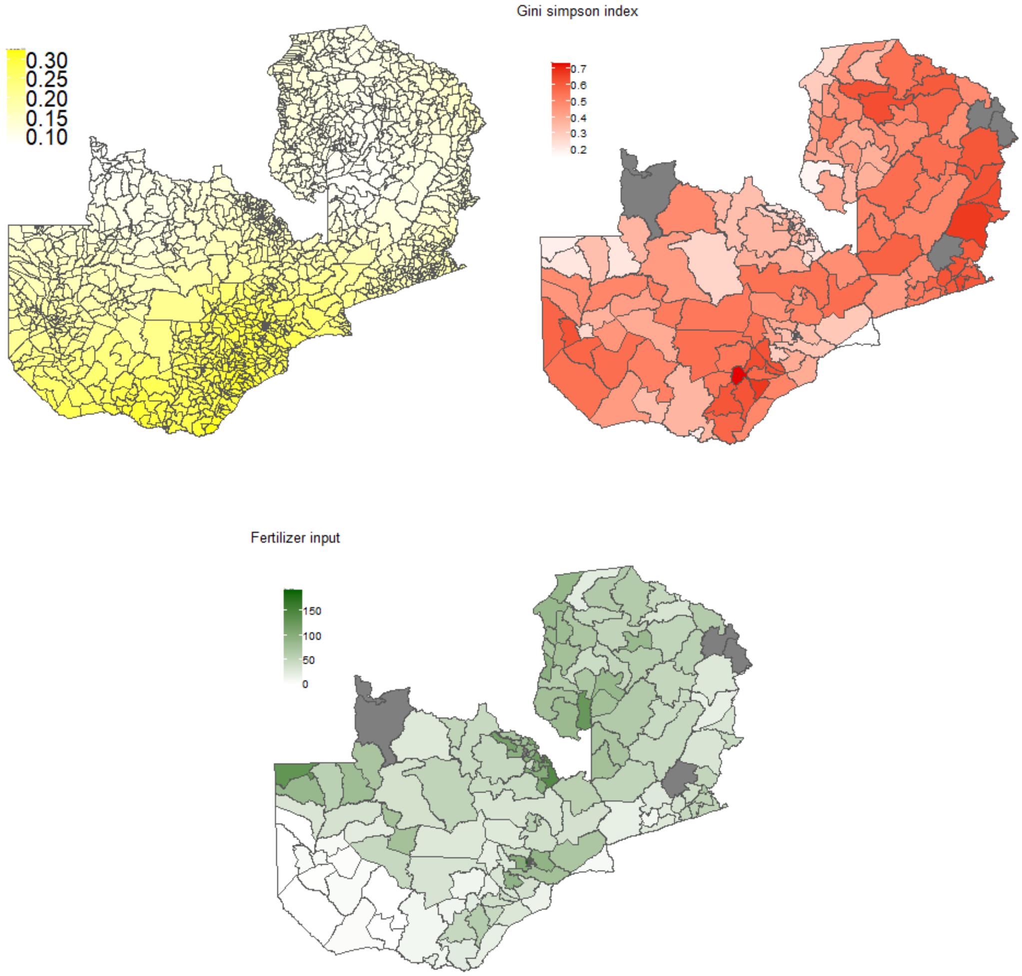

Figure 2: Rainfall variations, crop diversification, and fertilizer applications

Notes: The upper-left map plots the coefficient of variation of January and February rainfall for 59 years between the 1960/61 and 2018/19 cropping seasons by ward. The upper-right figure plots the average Gini-Simpson index of planted areas by crop as the crop diversification indicator for each constituency. The lower-middle map displays the average amount of fertilizer applied in kilograms per hectare for each constituency. Data are missing in the gray shaded areas.

risk index is unclear, the average amount of fertilizer applied tends to be high in the Copperbelt and Luapula provinces and low in the Western and Southern provinces, suggesting a negative association with climate risk. The regression analysis presented in the next section formally tests these observations.

5 Farmers' Risk Mitigation in Agricultural Decisions as a Response to Rainfall Risks

5.1 Specification

This section tests the response of farmers' risk management to location-specific rainfall risks by exploiting cross-sectional variations in interannual rainfall variability. To this end, we model farmers' agricultural decisions as follows:

$$Y_{iwp} = \beta_{CV} CV_{wp}^{12} + \mathbf{X}'_{iwp}\alpha + \mathbf{Z}'_{wp}\gamma + \phi_p + \varepsilon_{iwp} \tag{3}$$

where Y_{iwp} is the outcome of interest for household i in ward w of province p during the 2020/21 rainy season. The outcome variables include risk diversification measures and agricultural investments. For risk diversification, we analyze the number of crops, the Gini-Simpson index of the area cultivated for different crops, and the number of plots cultivated. Additionally, the regression analysis examines the cultivation of sorghum and millet, which are more drought-tolerant than maize, as another diversification measure (see Appendix Figure A.1). For agricultural investments, we investigate fertilizer applications per area planted, adoption of hybrid maize seeds, family labor, and the land utilization rate, defined as the ratio of areas cultivated to total areas owned.

The primary explanatory variable is CV_{wp}^{12}, the CV of January and February rainfall in ward w calculated using historical rainfall estimates from the WorldClim database for 59 years between the 1960/61 and 2018/19 cropping seasons. Therefore, the parameter of interest is β_{CV}, which captures the impact of location-specific rainfall risks on farmers' risk management and agricultural investments. The sign of β_{CV} depends on which outcome is used as the dependent variable. On the one hand, we expect $\beta_{CV} > 0$ in the regression with the risk diversification index as the outcome variable if Zambian households manage weather risks by diversifying crops and plot locations. On the other hand, the theory predicts $\beta_{CV} < 0$ in the regression with fertilizer applied per hectare as the outcome variable if farmers respond to climate risks by hesitating risky investments.

$\mathbf{X}_{iwp}$ is the vector of household-level controls, including the total size of land owned, subjective land soil quality, family size, and household head characteristics, such as gender, age, and educational attainment. In contrast, $\mathbf{Z}_{wp}$ represents the vector of ward-level controls, such as the average annual rainfall over 59 agricultural years (1960/61–2018/19) and objective soil quality measures.[19] We also add maximum and minimum temperatures and rainfall from November 2019 to February 2020 to capture transitory income shocks that may restrict farmers' agricultural decisions for the subsequent 2020/21 cropping season. By controlling for these immediate weather shocks, β_{CV} directly captures long-run behavioral reactions to rainfall risks. Finally, ϕ_p stands for province fixed effects, and ε_{iwp} is an error term.

The key identification assumption for the estimation of β_{CV} is that rainfall variability measured on historical data (CV_{wp}^{12}) should be uncorrelated with other determinants omitted from the right-hand side variables of regression equation (3). This exogeneity assumption of CV_{wp}^{12} is violated if systematic differences exist among wards that are correlated with CV_{w}^{12} and determine the average agricultural decisions in the area. For example, selective migration (e.g., when some households are more likely to migrate away from high-climate-risk regions) may induce this empirical concern. Section 7.2 examines the possibility of selective migration based on climate risk to ensure the validity of the identification assumption.

5.2 Results

Table 3 presents the estimation results for risk-diversification strategies in agricultural production. These results do not support farmers' management of rainfall risk through crop and plot location choices. These findings contradict the theoretical predictions of the model of risk-averse households in developing countries. Moreover, the null or even negative results for diversification outcomes are not consistent with prior empirical evidence from Zambia (e.g., Arslan et al., 2018). Columns (4) and (5) do not support planting millet and sorghum as risk-management strategies.

Table 4 presents the regression results for agricultural investments in fertilizers and seeds. Columns (1)–(3) suggest that farmers facing high rainfall risk apply less fertilizer, particularly basal fertilizers such as D-compounds, than their counterparts facing low rainfall risk. This result is also economically significant: A one standard deviation increase of CV by 0.07 reduces the fertilizer applied to the field by 14.9 kilograms per hectare, corresponding to approximately 13

[19] As for soil quality measures, we include estimated amounts of nitrogen, phosphorus, potassium, water holding capacity, and soil pH. These are obtained from the soil nutrient maps of sub-Saharan Africa available at the ISRIC–World Soil Information website.

Table 3: Rainfall risk and diversification

	# of Crop	Gini–Simpson index	# of Plot	Millet	Sorghum
CoV (Jan, Feb)	0.75	0.25	-0.72	0.47	-0.72*
	(1.61)	(0.35)	(1.92)	(0.54)	(0.43)
Average precipitation	0.10*	0.028**	0.28***	0.028*	-0.0095
	(0.053)	(0.011)	(0.060)	(0.015)	(0.0075)
Prec Nov, 19	-0.038	-0.055	-0.011	-0.058	-0.048
	(0.21)	(0.046)	(0.25)	(0.063)	(0.038)
Prec Dec, 19	0.36***	0.058*	0.25*	0.11**	0.030
	(0.13)	(0.033)	(0.15)	(0.043)	(0.044)
Prec Jan, 20	-0.34**	-0.038	-0.33**	-0.084*	-0.015
	(0.15)	(0.031)	(0.15)	(0.043)	(0.038)
Prec Feb, 20	-0.26**	-0.036	-0.39***	0.0065	-0.0021
	(0.13)	(0.031)	(0.13)	(0.034)	(0.032)
Temp (min) Nov, 19	-0.40***	-0.080***	-0.34***	0.0070	0.035
	(0.11)	(0.024)	(0.12)	(0.037)	(0.038)
Temp (max) Nov, 19	0.20**	0.018	0.091	-0.043	-0.015
	(0.092)	(0.019)	(0.11)	(0.026)	(0.017)
Temp (min) Dec, 19	0.26**	0.059**	0.27**	0.023	0.054
	(0.12)	(0.028)	(0.12)	(0.033)	(0.034)
Temp (max) Dec, 19	-0.17	-0.019	-0.33**	0.064*	0.042
	(0.15)	(0.031)	(0.16)	(0.037)	(0.028)
Temp (min) Jan, 20	0.18	0.029	0.015	0.039	-0.0030
	(0.16)	(0.036)	(0.17)	(0.052)	(0.035)
Temp (max) Jan, 20	0.16	0.047	0.25	0.0041	0.060
	(0.18)	(0.039)	(0.20)	(0.057)	(0.045)
Temp (min) Feb, 20	-0.072	-0.0053	0.10	-0.045	-0.066**
	(0.13)	(0.029)	(0.12)	(0.057)	(0.028)
Temp (max) Feb, 20	-0.15	-0.045	-0.040	-0.044	-0.091**
	(0.14)	(0.031)	(0.17)	(0.045)	(0.044)
Soil condition = Medium	-0.021	-0.000083	-0.018	-0.0086	-0.0046
	(0.041)	(0.0083)	(0.042)	(0.0054)	(0.0038)
Soil condition = High	0.12**	0.024**	0.17***	0.0080	-0.00061
	(0.051)	(0.010)	(0.051)	(0.0085)	(0.0050)
Total Nitrogen (ppm)	-0.00016	-0.000044	-0.000052	-0.00018	-0.00011
	(0.00051)	(0.00012)	(0.00052)	(0.00012)	(0.00010)
Total Phosphorus (ppm)	-0.0023***	-0.00054***	-0.0025***	-0.00043**	-0.00020
	(0.00061)	(0.00015)	(0.00079)	(0.00020)	(0.00014)
Extractable Potassium (ppm)	0.0011	0.00030	0.0011	0.00095**	0.0011***
	(0.0013)	(0.00029)	(0.0019)	(0.00046)	(0.00037)
Water holding capacity (mm)	0.0016	0.00030	0.0026	0.0012**	0.00098*
	(0.0022)	(0.00055)	(0.0023)	(0.00058)	(0.00057)
Soil pH (depth 0–5cm)	-0.042	-0.0035	0.0089	0.019**	0.011
	(0.032)	(0.0069)	(0.037)	(0.0084)	(0.0076)
Adj. R-Squared	0.39	0.27	0.39	0.19	0.14
Observations	12220	12220	12220	12220	12220

Notes: Robust standard errors clustered by district are reported in parentheses. Total land size, family size, head's gender, age, years of education, and province fixed effects are included but not reported. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 4: Rainfall risk and investments in fertilizers and seeds

	Basal (/ha)	Top (/ha)	Total (/ha)	Hybrid seed
CoV (Jan, Feb)	-125.5**	-87.6	-213.1*	1.46***
	(58.5)	(61.3)	(118.6)	(0.47)
Average precipitation	7.46***	9.05***	16.5***	0.028
	(2.09)	(2.15)	(4.18)	(0.017)
Prec Nov, 19	-13.4**	-11.0*	-24.4*	0.043
	(6.18)	(6.28)	(12.3)	(0.065)
Prec Dec, 19	16.0***	16.3***	32.2***	0.13***
	(5.25)	(5.30)	(10.5)	(0.040)
Prec Jan, 20	2.05	-0.80	1.25	0.076
	(6.16)	(6.14)	(12.2)	(0.049)
Prec Feb, 20	-20.5***	-23.2***	-43.7***	-0.16***
	(4.01)	(3.84)	(7.59)	(0.035)
Temp (min) Nov, 19	3.81	3.35	7.16	-0.013
	(3.90)	(4.20)	(8.02)	(0.032)
Temp (max) Nov, 19	-0.52	0.60	0.082	0.030
	(3.39)	(3.64)	(6.99)	(0.027)
Temp (min) Dec, 19	1.98	0.35	2.33	0.012
	(5.42)	(5.51)	(10.9)	(0.040)
Temp (max) Dec, 19	-1.84	-1.94	-3.77	0.0080
	(6.69)	(6.94)	(13.6)	(0.062)
Temp (min) Jan, 20	-17.1**	-14.8**	-32.0**	-0.050
	(6.49)	(6.84)	(13.2)	(0.058)
Temp (max) Jan, 20	1.69	1.26	2.95	0.019
	(6.00)	(5.96)	(11.9)	(0.051)
Temp (min) Feb, 20	7.18	7.68	14.9	-0.0068
	(5.29)	(5.48)	(10.7)	(0.048)
Temp (max) Feb, 20	-0.088	-1.64	-1.73	-0.047
	(6.32)	(6.45)	(12.7)	(0.060)
Soil condition = Medium	6.98***	7.48***	14.5***	0.066***
	(1.74)	(1.54)	(3.22)	(0.017)
Soil condition = High	5.61***	6.39***	12.0***	0.079***
	(1.93)	(1.82)	(3.69)	(0.020)
Total Nitrogen (ppm)	0.038*	0.030	0.068*	0.00023
	(0.019)	(0.020)	(0.039)	(0.00017)
Total Phosphorus (ppm)	-0.11***	-0.097***	-0.20***	-0.00044*
	(0.026)	(0.028)	(0.054)	(0.00022)
Extractable Potassium (ppm)	0.0093	0.011	0.020	0.00076
	(0.055)	(0.059)	(0.11)	(0.00046)
Water holding capacity (mm)	0.079	0.093	0.17	0.0022***
	(0.090)	(0.092)	(0.18)	(0.00078)
Soil pH (depth 0–5cm)	0.25	0.0014	0.25	-0.0026
	(1.26)	(1.31)	(2.56)	(0.0090)
Adj. R-Squared	0.26	0.26	0.27	0.24
Observations	12220	12220	12220	12220

Notes: Robust standard errors clustered by district are reported in parentheses. Total land size, family size, head's gender, age, years of education, and province fixed effects are included but not reported. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

percent of its standard deviation. As the regression controls for transitory income shocks by including rainfall amounts and temperature in the previous rainy season, these results capture farmers' long-term reactions to weather risks, rather than their short-term responses to climate shocks.

To examine household-level investments in maize seeds, we use the weighted average of the hybrid seed indicators across maize plots, with the area planted as the weight. The results in Column (4) indicate that farm households in regions with high rainfall variability are more likely to adopt hybrid maize seeds than those in regions with mild variability. For example, we anticipate an increase in the likelihood of planting hybrid seeds by ten percentage points if the rainfall risk index increases by one standard deviation, whereas the sample average is 74%. This result is counter-intuitive given that planting hybrid seeds is relatively costly. However, the following features of hybrid maize seeds provide meaningful interpretations of high hybrid seed adoption rates in regions with high rainfall risk. First, hybrid varieties are more drought-tolerant than traditional varieties. Hence, we can consider planting hybrid varieties a risk-mitigating technology. Second, hybrid maize varieties grow faster than the local varieties; thus, hybrid seeds can be replanted even after germination failure during the first planting. Therefore, planting hybrid maize can be a risk-coping method in the early stages of the agricultural season. Combining these observations with the lack of significant results for the seedling rate (not reported), we speculate that Zambian farmers consider planting hybrid maize a risk-hedging option rather than a risky investment option.

The discussion now turns to agricultural investments in family labor and land. Table 5 summarizes the regression results. The estimation results do not support the idea that agricultural households adjust their family labor in response to climate risks (Columns 1–3). By contrast, the operating rate of agricultural land in the rainy season is higher among farm households in high-climate-risk regions than among their counterparts in low-climate-risk regions, although the coefficient is marginally significant. Given that the average land operation rate in the sample is 91%, rainfall risks encourage farmers to make full use of accessible fields. The full use of agricultural land may be motivated by compensation for the loss of production from hesitant fertilizer applications because of uninsured rainfall risks.

In summary, the estimation results find no evidence that precipitation risks promote crop and plot diversification strategies or the adoption of drought-tolerant crops, such as millet and sorghum. Instead, farmers respond to climate risks by reducing risky investments in fertilizers at the cost of high returns, making full use of accessible agricultural lands as a less costly investment to com-

Table 5: Rainfall risk, family labor, and land utilization

	Labor	Female labor	Male labor	Area Planted/Area Field
CoV (Jan, Feb)	0.99	0.16	0.83	0.51
	(1.10)	(0.72)	(0.79)	(0.34)
Average precipitation	0.038	0.028	0.010	-0.034***
	(0.038)	(0.025)	(0.023)	(0.0083)
Prec Nov, 19	0.31*	0.092	0.22**	0.00017
	(0.16)	(0.096)	(0.096)	(0.030)
Prec Dec, 19	-0.075	-0.048	-0.027	0.035*
	(0.099)	(0.060)	(0.061)	(0.020)
Prec Jan, 20	0.15*	0.14**	0.014	-0.013
	(0.091)	(0.057)	(0.063)	(0.020)
Prec Feb, 20	-0.15*	-0.054	-0.099*	0.025
	(0.087)	(0.057)	(0.057)	(0.020)
Temp (min) Nov, 19	-0.074	-0.0081	-0.066	-0.027
	(0.077)	(0.041)	(0.055)	(0.019)
Temp (max) Nov, 19	0.098*	-0.012	0.11***	0.049***
	(0.055)	(0.036)	(0.036)	(0.016)
Temp (min) Dec, 19	-0.014	-0.053	0.039	0.0046
	(0.083)	(0.054)	(0.050)	(0.020)
Temp (max) Dec, 19	-0.19**	-0.061	-0.13**	0.0098
	(0.092)	(0.058)	(0.056)	(0.022)
Temp (min) Jan, 20	-0.30***	-0.055	-0.24***	0.064**
	(0.11)	(0.063)	(0.071)	(0.030)
Temp (max) Jan, 20	-0.090	-0.031	-0.059	-0.032
	(0.13)	(0.070)	(0.083)	(0.024)
Temp (min) Feb, 20	0.38***	0.100	0.28***	-0.062***
	(0.11)	(0.062)	(0.070)	(0.023)
Temp (max) Feb, 20	0.17	0.10	0.069	-0.018
	(0.11)	(0.065)	(0.076)	(0.022)
Soil condition = Medium	0.011	-0.0037	0.014	0.0022
	(0.047)	(0.026)	(0.031)	(0.0047)
Soil condition = High	-0.0042	0.015	-0.019	-0.0035
	(0.055)	(0.031)	(0.036)	(0.0065)
Total Nitrogen (ppm)	-0.00031	-0.00016	-0.00015	-0.000039
	(0.00043)	(0.00027)	(0.00022)	(0.00011)
Total Phosphorus (ppm)	0.00063	0.00014	0.00049	-0.00011
	(0.00057)	(0.00035)	(0.00030)	(0.00019)
Extractable Potassium (ppm)	0.00043	0.00084	-0.00041	0.00036
	(0.0011)	(0.00078)	(0.00067)	(0.00037)
Water holding capacity (mm)	-0.0015	-0.00089	-0.00063	0.00016
	(0.0018)	(0.0012)	(0.0012)	(0.00042)
Soil pH (depth 0–5cm)	-0.015	0.0040	-0.019	-0.0075
	(0.018)	(0.011)	(0.012)	(0.0050)
Adj. R-Squared	0.54	0.37	0.38	0.43
Observations	12220	12220	12220	12220

Notes: Robust standard errors clustered by district are reported in parentheses. Total land size, family size, head's gender, age, years of education, and province fixed effects are included but not reported. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

pensate for the loss, and adopting hybrid maize seeds as risk management and coping strategies.

The natural question is: Why do Zambian farmers fail to pursue risk diversification through crop choices? One possible explanation for this finding is the regional heterogeneity. For example, Arslan et al. (2018) show that in Zambia, farmers in relatively heavy precipitation regions diversify their crops, while farmers in other regions diversify their income sources and livestock portfolios. Thus, households may diversify their income risks owing to climate variability in other dimensions. Moreover, the agronomic growing conditions and requirements are different for each crop, which innately restricts the choice of crop variety. The ineffectiveness of crop diversification in practice, misunderstanding of the effectiveness of diversification among small-scale farmers, and strong preferences for maize as a food crop also provide alternative potential explanations. Data constraints did not allow us to examine the reasons for the negligible crop diversification in response to climate risks. Determining the reasons for this phenomenon is a promising avenue for future research.

Instead, our discussion raises a different question. Our empirical results reveal that rainfall risk significantly affects investment decisions in farming, suggesting that it affects agricultural productivity. By quantifying the productivity impacts of rainfall risks, we can determine the cost of climate risks and the potential benefits of mitigating them. Thus, the extent to which behavioral changes induced by climate risk miss agricultural outputs is an attractive question for policymakers and governments. The mediation analysis, described in the next section, addresses this question.

6 Mediation Analysis

6.1 Estimating Mediating Effects

In this section, we examine how risk avoidance in agricultural decisions affects the final maize productivity by estimating the mediating effects. The mediation analysis focuses on hybrid seed adoption and fertilizer application as risk-induced responses in the form of agricultural investments.

One natural estimand for examining mediation effects is the average natural directed effect (ANDE), which is the impact of precipitation risk on yield conditional on seed choice or fertilizer application. However, estimating the ANDE without bias is empirically challenging. To examine this, we consider a naïve regression model to estimate the effects of treatment D (CV of January and February rainfall in our case) on outcome Y (maize yield) conditional on mediator M (hybrid seeds and fertilizer):

$$Y_i = \beta_0 + \beta_1 D_i + \beta_2 M_i + \beta_3 X_i + \beta_4 Z_i + \epsilon_i \tag{4}$$

where X_i and Z_i represent the pretreatment and intermediate confounders, respectively (Acharya et al., 2016). The estimation problem to identify the regression estimator for β_1 is that it may contain the intermediate variable bias, selection bias that can arise from the inclusion of variables affected by treatment as controls in the regression model.[20] Under this potential bias, the OLS estimator for β_1 representing the ANDE of treatment cannot be consistent and unbiased.

Another parameter of interest is the average natural indirect effect (ANIE), which captures the impact of the subsequent change in mediator M induced by the change in treatment D while fixing the effect of the treatment. Denoting $Y(d, m)$ as the potential outcome for the realized treatment $D = d$ and mediator $M = m$, the ANIE can be represented as:

$$ANIE(d, d') = E[Y_i(d, M_i(d)) - Y_i(d, M_i(d'))]$$

However, the ANIE is not identified in the presence of intermediate confounders, that are affected by the intervention and affect the outcome. To get around the issue, we indirectly estimate the ANIE as a residual by exploiting the fact that the average treatment effect (ATE) can be decomposed into a linear sum of the average conditional direct effect (ACDE), the ANIE, and interaction effects. ACDE is the average causal effect of the treatment when the mediator variables are fixed for all observations at a specific value (Acharya et al., 2016). For example, the ACDE of the change in treatment from d to d' is represented as:

$$ACDE(d, d', m) = E[Y_i(d, m) - Y_i(d', m)],$$

Conversely, the interaction effect is the average extent to which the direct effect differs according to the mediators. Mathematically, the interaction effect is defined as:

$$E[M(d')[CDE(d, d', m) - CDE(d, d', m)]]$$

[20] The two sources of the intermediate variable bias are the classical omitted variable bias and the bias from blocking the path of $D \to Z \to Y$ due to the inclusion of intermediate confounders Z.

With these definitions, the ANIE estimation involves two steps. After estimating the ATE and ACDE in the first step, subtracting the ACDE from the ATE provides the estimator for the ANIE plus the interaction effect, which is an approximation of the indirect effect when the interaction effect is not substantial.

In the first step of the estimation procedure, the consistency of a regression-based estimator for the ACDE requires the following two identification conditions.[21] The first assumption is called sequential unconfoundedness. Formally, the assumption of sequential unconfoundedness can be expressed as:

Assumption 1 *Sequential unconfoundedness*

$$Y_i(d, m), M_i(d) \perp\!\!\!\perp D_i | X_i,$$
$$Y_i(d, m) \perp\!\!\!\perp M_i | D_i, X_i, Z_i$$

Assumption 1 does not allow for two types of omitted variables: Those for the effect of treatment on the outcome, conditional on the pretreatment confounders, and those for the effect of the mediator on the outcome, conditional on the treatment, pretreatment confounders, and intermediate confounders. This condition ensures a separate estimation of the impact of the treatment and the mediator on the outcome.

The second assumption for ACDE identification is the absence of intermediate interactions. This assumption can be stated as follows:

Assumption 2 *No intermediate interaction*

$$E[Y_i(d, m) - Y_i(d, m') | D_i, X_i, Z_i] = E[Y_i(d, m) - Y_i(d, m') | D_i, X_i]$$

Assumption 2 requires the effect of the mediator on the outcome and the intermediate confounders to be independent. These two assumptions are necessary for an unbiased mediation analysis.

[21] Semi-parametric and non-parametric estimators do not require the assumption of no intermediate interactions (Assumption 2). However, these alternative estimators are unsuitable when the treatment and mediator variables are continuous, as is the case in our empirical setting. See Acharya et al. (2016) for further details.

We apply the two-stage derivation of the regression-based estimator to our empirical setting. This empirical exercise uses plot-level CFS data because plot characteristics, which are important determinants of maize yield, can be controlled for in the regression analysis. The first-stage regression estimates the impact of the mediators (i.e., hybrid seed adoption and fertilizer application) on the outcomes (i.e., maize yield). Specifically, we run the following regression equation in the first stage:

$$\log Yield_{liwp} = \beta_1 CV^{12}_{wp} + M'_{liwp}\beta_2 + \mathbf{X}'\beta_3 + \mathbf{Z}'\beta_4 + \delta_p + \epsilon_{liwp} \tag{5}$$

where $\log Yield_{liwp}$ is the logarithm of (expected) maize yield in plot l of household i in ward w and M_{liwp} is a vector of the dummy taking one if household i plants hybrid maize seeds in plot l and the amount of fertilizer applied per hectare to plot l by household i. $\mathbf{X}$ contains pre-determined covariates at the plot (soil conditions), household (e.g., sex and age of household i's head), and ward levels (monthly rainfall amounts and temperatures from November 2020 to February 2021 to control for current productivity shocks). $\mathbf{Z}$ contains post-determined covariates including the family size of household i in ward w. δ_p stands for fixed effects for province p, and ϵ_{liwp} is an error term.

In the second stage, we regress the demediated outcome as $\log \widetilde{Yield}_{liwp} = \log Yield_{liwp} - M'_{liwp}\hat{\beta}_2$ on the treatment and controls:

$$\log \widetilde{Yield}_{liwp} = \alpha_1 CV^{12}_{wp} + \mathbf{X}'\alpha_3 + \mathbf{Z}'\alpha_4 + \delta_p + \varepsilon_{liwp} \tag{6}$$

In this regression model, the coefficient α_1 represents the ACDE of the rainfall risk CV^{12}_{w}. As the standard errors in the second regression are biased owing to the estimation error in the first-stage regression, we use standard non-parametric bootstrap methods in both stages.

The credibility of the mediation analysis depends on the specification of Equation (5) and the validity of the identification assumptions. In Section 7, we test the robustness of the mediation analysis results by conducting a sensitivity analysis of the violations of Assumption 1.

6.2 Results

Table 6 summarizes the mediation analysis results.[22] In the first column, the estimated ATE is -1.40, suggesting that a one-standard-deviation increase in the rainfall risk index by 0.07 dimin-

[22] Appendix Table A5 presents summary statistics for empirical variables used for the estimation.

Table 6: Estimated ATE and ACDEs

	Total	Fertilizer	Hybrid seed
CoV (Jan, Feb)	-1.396***	-.858**	-2.314***
	(0.468)	(0.406)	(0.422)
Observations	11429	11429	11429

Notes: Column 1 reports the average treatment effect of the coefficient of variation for the rainfall in January and February. Robust standard errors clustered by household are reported in the parenthesis. Columns 2 and 3 reports the average conditional direct effects of the coefficient of variation for the rainfall in January and February, conditional on fertilizer (Column 2) and seed inputs (Column 3). Non-parametric bootstrap standard errors based on 1,000 replications are reported in parentheses.

ishes maize yields by 9.3 ($= \exp(-1.396 \times 0.07) - 1) \times 100$) percent. The long-term rainfall risk is the sole cause of this 9.3% gap in maize yield as the regression controls for soil conditions and weather-related productivity shocks. Given that a non-negligible number of Zambian farmers live near subsistence levels, the estimated risk impacts on staple food production are significant in absolute terms. The key observation is that the direct effect of precipitation risk on yield may come from additional factors that affect productivity other than soil conditions, rainfall conditions in that year, and endogenous risk management strategies.

The second and third columns present the ACDE of historical rainfall variations conditional on fertilizer and seed inputs. For example, the estimated coefficient in the second column represents the effects of rainfall risk when the amount of fertilizer applied is fixed to the empirical sample average. Conversely, the third column shows the impact of rainfall risk when no adoption of hybrid maize seeds is assumed. The estimation results imply that maize yields decrease by 5.8% after a one-standard-deviation increase in rainfall risk when fertilizer inputs are conditioned at the sample average. The same increase in rainfall risk depresses the maize yield by 15.0% when hybrid maize is not planted. In other words, if farmers do not use fertilizers in both the high- and low-risk regions, the impact of rainfall risk on maize productivity decreases by approximately 38.5% ($\frac{1.396-0.858}{1.396} \times 100$) relative to the ATE. In contrast, if farmers use hybrid seeds in both high- and low-risk regions, the treatment effect of rainfall risk on maize productivity increases by approximately

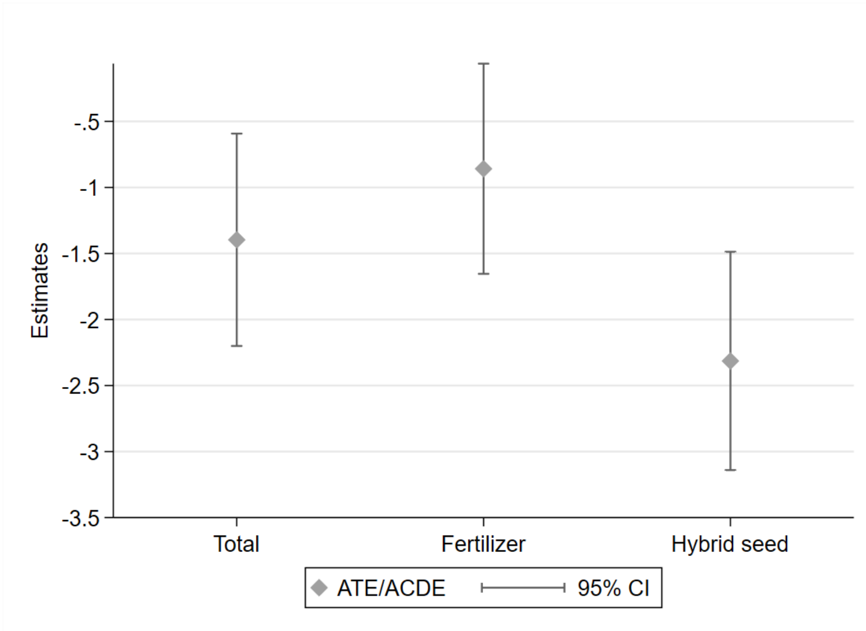

Figure 3: ATE and ACDEs of rainfall risks

Notes: This figure illustrates the point estimates and 95% confidence intervals for the average treatment effect and average conditional treatment effects.

65.8% ($\frac{2.314-1.396}{1.396} \times 100$) relative to the ATE. Figure 3 graphically exhibits the mediation analysis results.

7 Robustness Check

This section confirms the robustness of the main empirical results by altering the proxy for rainfall risk, discussing the possibility of selective migration as a source of endogenous climate risk, and verifying the sensitivity of the mediation analysis results to violations of the sequential unconfoundedness assumption.

7.1 Alternative Definition of Precipitation Risk

The first concern is the misspecification of precipitation risk measures defined in Section 4. Instead of the coefficient of variation for the rainfall in January and February over 59 years, we check the robustness of the main results to alternative definitions of precipitation risk. Particularly, we use (1) rainfall during the entire agricultural season between November and April or (2) rainfall during the planting and weeding seasons between November and February, and then calculate the coefficient of variation for each rainfall. The regression results in Appendix Tables A6–A11 are similar to the main results, although some coefficients lose statistical significance.

7.2 Endogeneity in Climate Risk: Migration

Another empirical concern is the potential endogeneity of the precipitation risk. Selective migration is a potential source of endogeneity. A possible scenario is that if better-off households migrate away from high-risk regions, only resource-poor households will remain concentrated in disadvantaged regions. This scenario systematically differentiates between low- and high-rainfall-risk regions in terms of the production resources that determine agricultural decisions.

As a simple empirical test for this possibility, we examine if the change in the regional average of asset levels is correlated with our rainfall risk index using data from the Census of Population and Housing in 2000 and 2010.[23] To construct the asset index, the principal component analysis calculates the asset score based on the ownership of durable goods for each household, and then we aggregate them at the constituency level.[24,25]

While the left panel in Figure 4 plots the change in the ranking of asset scores at the constituency level, the right panel presents a scatter plot between the rank change and our rainfall risk index, that is, the CV of the January and February rainfall based on historical rainfall data at the constituency level. We find no correlation between ranking changes based on asset scores and climate risk. Thus, this empirical exercise does not provide supporting evidence for selective migration based on production resources.

[23]The 10% sample microdata of the Zambian Census of Population and Housing are available at https://international.ipums.org/international/.

[24]For the 2000 Census data, we use the following durable goods as components of the asset index: Refrigerators, radios, kitchens, motorcycles, motor vehicles, telephones, and roof materials. In addition to these durable assets, we add the following to the list of score components when calculating the asset index using 2010 Census data: Televisions, bicycles, Internet facilities, computers, and mobile phones. Although the sources of the asset index differ across census years, using the same set of durable assets is not necessary because we compare the rankings of the constituency based on asset scores rather than comparing the asset indexes per se.

[25]Constituencies are the finest geographic units available in the census for both survey years.

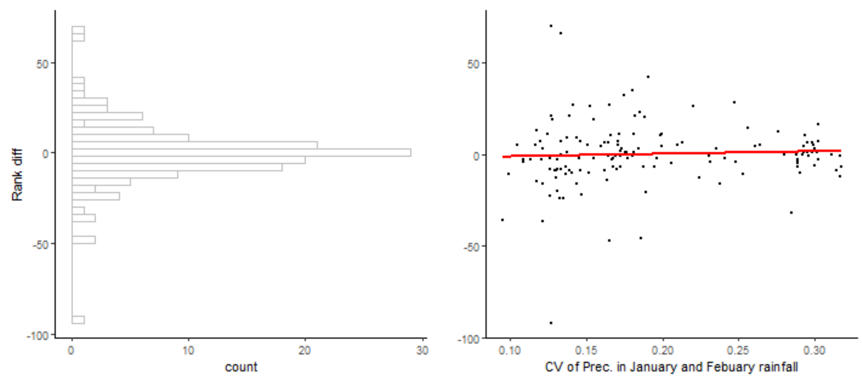

Figure 4: Asset scores and climate risk by constituency, 2000 and 2010

Notes: The left figure is a histogram of the change in the ranking of the asset index based on the first principal components of durable goods from 2000 to 2010. The figure on the right depicts the relationship between the change in the ranking of the asset index and the coefficient of variation for the rainfall in January and February over the past 59 years.

7.3 Sensitivity Analysis of Mediation Analysis

The credibility of the ADCE estimation depends on the validity of Assumptions 1 and 2 and the specifications of regression equation (5). Among these, we can assess the violations of Assumption 1, which is *sequential unconfoundedness*. This assumption requires that the treatment assignment of rainfall risks CV_w^{12} should be uncorrelated with potential outcomes and potential mediators after conditioning on pretreatment covariates, and that mediators should be uncorrelated with potential outcomes after controlling for treatment, pretreatment confounders, and intermediate confounders. In other words, these conditions allow us to consistently estimate the effects of rainfall risk, fertilizer input, and hybrid seed utilization on maize yield using OLS. However, because input choices, such as hybrid seeds and chemical fertilizers, are part of complicated household decisions, some unobservables may violate the latter condition.

Acharya et al. (2016) propose a sensitivity analysis to violate sequential unconfoundedness. The sensitivity analysis is based on the observation that bias arises from the correlation between the error terms in equations (5) and (7).[26] Therefore, we can characterize the violation of the sequential unconfoundedness assumption by estimating the ACDE for different hypothetical values of the correlation between the mediator and outcome errors.

[26] Appendix Section A.4 presents the bias form.

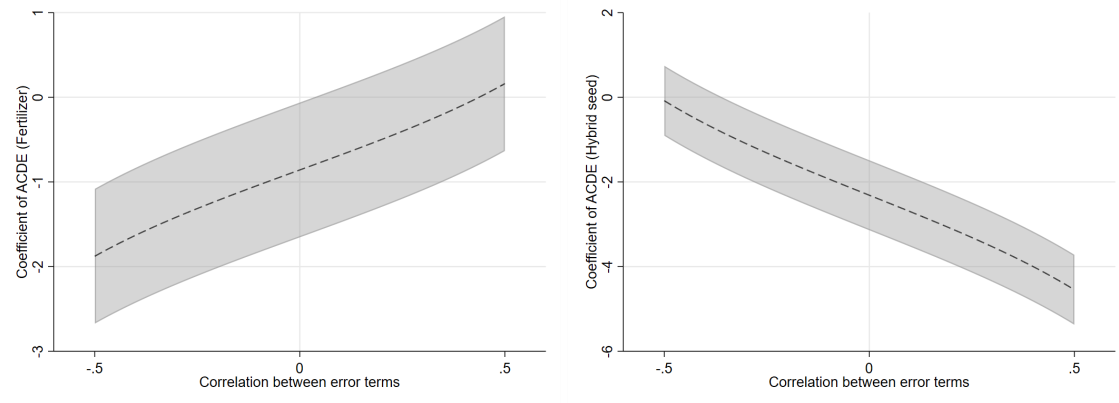

Figure 5: Estimated ACDEs for different correlations between error terms

Notes: This figure depicts the relationship between the correlation of error terms and ACDEs, conditional on fertilizer input (left) and hybrid seed usage (right). The gray shaded regions represent the 95% confidence interval of the ACDE. The construction of confidence intervals does not consider the sample uncertainty of the bias.

Figure 5 depicts the estimated ACDE under different correlation-level assumptions. If unmeasured factors make both fertilizer applications and maize yields positively correlated (considered more plausible in this case), the impact of climate risks on yields at a constant level of fertilizer input will become smaller in absolute terms than when there is no such bias. The convergence of the estimated ACDE to zero as the positive correlation between fertilizer application and yield becomes strong implies that climate adaptations are made primarily through adjustments in the agricultural input portfolio, particularly when agricultural decisions are interconnected and complementary. Thus, this sensitivity analysis result highlights the significant productivity implications of household responses to climate risk through fertilizer investment. Conversely, the sensitivity analysis results for the ACDE, conditional on the lack of hybrid maize seeds in the right panel of Figure 5, suggest that accounting only for maize seed choice leaves significant and independent climate risk impacts unexplained.

8 Conclusion

Active debates in the climate-policy arena require a comprehensive understanding of farmers' responses to weather risks in developing countries. This study contributes to the literature by providing micro-level evidence for risk management in agricultural production among Zambian farmers. Our empirical results found no evidence that rainfall risks promote crop and plot diversification strategies or the adoption of drought-tolerant crops, such as sorghum and millet. Instead, they respond to climate risks by reducing fertilizer application as a risky investment, expanding planted agricultural land as a less costly investment, and adopting hybrid maize seeds as risk management and coping strategies. We also found that, after accounting for soil conditions and recent climate-related productivity shocks, the maize yield gap is approximately 9% when the difference in our climate risk index equals one standard deviation. Our mediation analysis focused on fertilizer application and hybrid seed adoption as essential pathways through which climate risk affects household maize production. Although the results indicate that adopting hybrid maize seeds generates yield-enhancing effects, their favorable impacts are attenuated by the negative impacts of underinvestment in chemical fertilizers in response to rainfall risks. Overall, the empirical evidence suggests that household-level climate adaptations are made primarily through adjustments in the agricultural input portfolio rather than through risk diversification strategies in Zambia.

We conclude the paper by suggesting two promising avenues for future research. First, our finding of no diversification in response to rainfall risk raises the question of why Zambian farmers fail to pursue risk diversification through their crop choices. Future research should provide rational explanations for this empirical puzzle and propose policy interventions to relax these constraints. Second, this study did not consider heterogeneity in responses to climate risks. For example, access to off-farm activities may cushion the impact of climate risk on farm income, allowing farmers to make different agricultural decisions. This potential interplay of risk management strategies suggests the importance of identifying cost-effective ways to control the consequences of climate risks faced by smallholders. The data constraints prevented us from exploring these critical possibilities. Incorporating other income-generating activities into the empirical analysis along with further data collection will enrich our understanding of farmers' risk management in developing countries.

References

Acharya, A., Blackwell, M., and Sen, M. (2016). Explaining causal findings without bias: Detecting and assessing direct effects. *American Political Science Review*, 110(3):512–529.

Alem, Y., Bezabih, M., Kassie, M., and Zikhali, P. (2010). Does fertilizer use respond to rainfall variability? Panel data evidence from Ethiopia. *Agricultural Economics*, 41(2):165–175.

Aragón, F. M., Oteiza, F., and Rud, J. P. (2021). Climate change and agriculture: Subsistence farmers' response to extreme heat. *American Economic Journal: Economic Policy*, 13(1):1–35.

Arslan, A., Cavatassi, R., Alfani, F., Mccarthy, N., Lipper, L., and Kokwe, M. (2018). Diversification under climate variability as part of a CSA strategy in rural Zambia. *The Journal of Development Studies*, 54(3):457–480.

Boucher, S. R., Carter, M. R., Flatnes, J. E., Lybbert, T. J., Malacarne, J. G., Marenya, P., and Paul, L. A. (2021). Bundling genetic and financial technologies for more resilient and productive small-scale agriculture. Technical report, National Bureau of Economic Research.

Chen, S., Chen, X., and Xu, J. (2016). Impacts of climate change on agriculture: Evidence from China. *Journal of Environmental Economics and Management*, 76:105–124.

Chen, S. and Gong, B. (2021). Response and adaptation of agriculture to climate change: Evidence from China. *Journal of Development Economics*, 148:102557.

Cui, X. and Xie, W. (2022). Adapting agriculture to climate change through growing season adjustments: Evidence from corn in China. *American Journal of Agricultural Economics*, 104(1):249–272.

Dercon, S. (1996). Risk, crop choice, and savings: Evidence from Tanzania. *Economic Development and Cultural Change*, 44(3):485–513.

Emerick, K., De Janvry, A., Sadoulet, E., and Dar, M. H. (2016). Technological innovations, downside risk, and the modernization of agriculture. *American Economic Review*, 106(6):1537–61.

Fafchamps, M., Udry, C., and Czukas, K. (1998). Drought and saving in West Africa: are livestock a buffer stock? *Journal of Development Economics*, 55(2):273–305.

Fick, S. E. and Hijmans, R. J. (2017). WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology*, 37(12):4302–4315.

Harris, I., Osborn, T. J., Jones, P., and Lister, D. (2020). Version 4 of the CRU TS monthly high-resolution gridded multivariate climate dataset. *Scientific Data*, 7(1):109.

Ito, T. and Kurosaki, T. (2009). Weather risk, wages in kind, and the off-farm labor supply of agricultural households in a developing country. *American Journal of Agricultural Economics*, 91(3):697–710.

Karlan, D., Osei, R., Osei-Akoto, I., and Udry, C. (2014). Agricultural decisions after relaxing credit and risk constraints. *Quarterly Journal of Economics*, 129(2):597–652.

Kurukulasuriya, P., Mendelsohn, R., Hassan, R., Benhin, J., Deressa, T., Diop, M., Eid, H. M., Fosu, K. Y., Gbetibouo, G., Jain, S., et al. (2006). Will African agriculture survive climate change? *World Bank Economic Review*, 20(3):367–388.

Liu, M., Shamdasani, Y., and Taraz, V. (2023). Climate change and labor reallocation: Evidence from six decades of the Indian Census. *American Economic Journal: Economic Policy*, 15(2):395–423.

Lobell, D. B., Bänziger, M., Magorokosho, C., and Vivek, B. (2011). Nonlinear heat effects on African maize as evidenced by historical yield trials. *Nature Climate Change*, 1(1):42–45.

Maidment, R. I., Grimes, D., Allan, R. P., Tarnavsky, E., Stringer, M., Hewison, T., Roebeling, R., and Black, E. (2014). The 30 year tamsat african rainfall climatology and time series (tarcat) data set. *Journal of Geophysical Research: Atmospheres*, 119(18):10–619.

Maidment, R. I., Grimes, D., Black, E., Tarnavsky, E., Young, M., Greatrex, H., Allan, R. P., Stein, T., Nkonde, E., Senkunda, S., et al. (2017). A new, long-term daily satellite-based rainfall dataset for operational monitoring in Africa. *Scientific Data*, 4(1):1–19.

Mendelsohn, R. and Dinar, A. (2003). Climate, water, and agriculture. *Land Economics*, 79(3):328–341.

Ministry of Agriculture and Ministry of Fisheries and Livestock (2016). Second national agricultural policy. Technical report.

Miura, K., Kanno, H., and Sakurai, T. (2012). Shock and livestock transactions in rural Zambia: a re-examination of the buffer stock hypothesis. *Japanese Journal of Rural Economics*, 14:20–34.

Morduch, J. (1995). Income smoothing and consumption smoothing. *Journal of Economic Perspectives*, 9(3):103–114.

Newbery, D. M. (1991). Agricultural Institutions for Insurance and Stabilization. In *The Economic Theory of Agrarian Institutions*. Oxford University Press.

Taraz, V. (2018). Can farmers adapt to higher temperatures? Evidence from India. *World Development*, 112:205–219.

Tarnavsky, E., Grimes, D., Maidment, R., Black, E., Allan, R. P., Stringer, M., Chadwick, R., and Kayitakire, F. (2014). Extension of the tamsat satellite-based rainfall monitoring over Africa and from 1983 to present. *Journal of Applied Meteorology and Climatology*, 53(12):2805–2822.

Waldman, K. B., Blekking, J. P., Attari, S. Z., and Evans, T. P. (2017). Maize seed choice and perceptions of climate variability among smallholder farmers. *Global Environmental Change*, 47:51–63.

Welch, J. R., Vincent, J. R., Auffhammer, M., Moya, P. F., Dobermann, A., and Dawe, D. (2010). Rice yields in tropical/subtropical Asia exhibit large but opposing sensitivities to minimum and maximum temperatures. *Proceedings of the National Academy of Sciences*, 107(33):14562–14567.

World Bank (2021). Employment in agriculture (% of total employment). Data retrieved from World Development Indicators, https://data.worldbank.org/indicator/SL.AGR.EMPL.ZS?locations=ZMM.

World Bank (2022a). Agriculture, forestry, and fishing, value added (% of GDP). Data retrieved from World Development Indicators, https://data.worldbank.org/indicator/NV.AGR.TOTL.ZS?locations=ZM.

World Bank (2022b). Rural population (% of total population). Data retrieved from World Development Indicators, https://data.worldbank.org/indicator/SP.RUR.TOTL.ZS?locations=ZM.

ZVAC (2015). In-depth vulnerability and needs assessment report. Zambia Vulnerability Assessment Committee. June, Lusaka, Zambia.

A Appendix

A.1 Appendix Figure

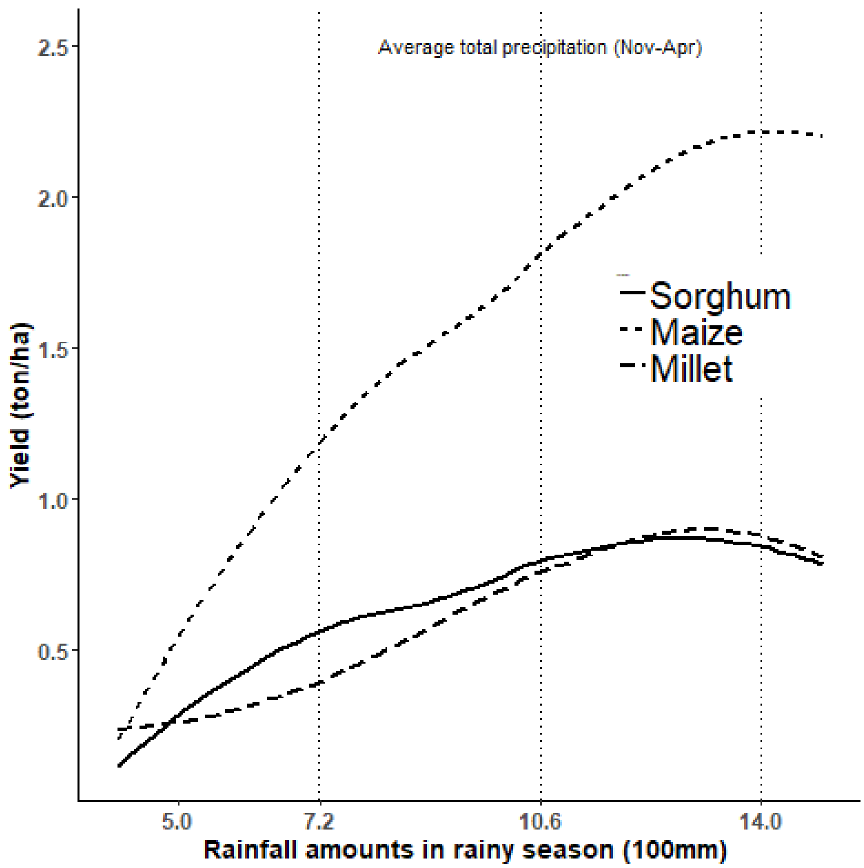

Figure A1: Rainfall and crop yields

Notes: This figure depicts the district-level relationship between rainfall from November to February during the rainy season and average crop yields. Agricultural statistics aggregated at the district level from the CFS are used for the estimation.

A.2 Appendix Tables

Table A1: Summary statistics: Variables used for the estimation of the maize production function at the district level

Variable	Mean	Std. Dev.	Min	Max	Obs
Yield (1000kg/ha)	1.70	0.99	0.00	12.00	1734
Prec Jan	0.24	0.07	0.07	0.48	1734
Prec Feb	0.20	0.06	0.03	0.42	1734
Prec Mar	0.17	0.07	0.01	0.49	1734
Prec Apr	0.05	0.04	0.00	0.21	1734
Prec May	0.01	0.01	0.00	0.14	1734
Prec Jun	0.00	0.00	0.00	0.01	1734
Prec Oct	0.02	0.02	0.00	0.12	1734
Prec Nov	0.12	0.05	0.02	0.34	1734
Prec Dec	0.23	0.07	0.06	0.42	1734
Prec Weed	0.46	0.15	0.12	1.70	1734
Prec Plant	0.37	0.13	0.13	1.30	1734

Notes: This table presents the summary statistics for the variables used to estimate the maize production function at the district level. Maize yield was computed using agricultural statistics aggregated at the district level from the CFS. The other variables were historical precipitation estimates aggregated at the district level using the WorldClim database.

Table A2: Robustness to adding time fixed effects

	All	Season	Season interaction
Prec Nov	0.66		
	(0.72)		
Prec Dec	-0.58		
	(0.48)		
Prec Jan	0.71*		
	(0.40)		
Prec Feb	2.71***		
	(0.52)		
Prec Plant		-0.44**	0.16
		(0.21)	(0.32)
Prec Weed		0.67***	1.24***
		(0.19)	(0.28)
Prec Plant × Prec Weed			-0.98**
			(0.39)
Prec Mar	0.70	1.53***	1.36**
	(0.48)	(0.53)	(0.52)
Prec Apr	-3.43***	-3.31***	-3.23***
	(0.97)	(1.09)	(1.08)
Adj. R-Squared	0.54	0.54	0.54
Observations	1734	1734	1734

Notes: Robust standard errors clustered by district are reported in parentheses. Province fixed effects and year dummies are included, but not reported. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A3: Robustness to using district fixed effects

	All	Season	Season interaction
Prec Nov	0.88*		
	(0.46)		
Prec Dec	0.62		
	(0.51)		
Prec Jan	0.34		
	(0.36)		
Prec Feb	2.93***		
	(0.32)		
Prec Plant		-0.087	0.99***
		(0.24)	(0.30)
Prec Weed		0.93***	1.91***
		(0.19)	(0.21)
Prec Plant × Prec Weed			-1.79***
			(0.36)
Prec Mar	-0.16	0.46	0.30
	(0.33)	(0.33)	(0.33)
Prec Apr	-0.73	-0.50	-0.51
	(0.66)	(0.60)	(0.62)
Linear trend in year	-0.057***	-0.047***	-0.047***
	(0.014)	(0.014)	(0.014)
Square trend in year	0.0028***	0.0028***	0.0027***
	(0.00047)	(0.00047)	(0.00048)
Adj. R-Squared	0.49	0.48	0.48
Observations	1734	1734	1734

Notes: Robust standard errors clustered by district are reported in parentheses. District dummies are included, but not reported. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A4: Summary statistics: Control variables

Variable	Mean	Std. Dev.	Min	Max	Obs
Total land size (ha)	2.90	2.70	0.01	32.0	12220
Family size	5.90	2.70	1.00	24.0	12220
Educ years, head	7.60	3.00	1.00	15.0	12220
Age, head	46.00	14.00	13.00	99.0	12220
Female head dummy	0.17	0.38	0.00	1.0	12220
Average precipitation	10.00	2.00	5.50	15.0	12220
Prec Jan, 20	2.20	0.36	1.20	3.1	12220
Prec Feb, 20	2.50	0.47	1.50	4.0	12220
Prec Nov, 19	1.00	0.45	0.21	2.2	12220
Prec Dec, 19	2.40	0.35	1.60	3.8	12220
Temp (max) Jan, 20	27.00	1.80	17.00	42.0	12220
Temp (max) Feb, 20	28.00	1.90	18.00	43.0	12220
Temp (max) Nov, 19	31.00	2.40	19.00	48.0	12220
Temp (max) Dec, 19	27.00	1.80	18.00	42.0	12220
Temp (min) Jan, 20	17.00	1.60	9.90	26.0	12220
Temp (min) Feb, 20	17.00	1.50	10.00	26.0	12220
Temp (min) Nov, 19	18.00	1.90	10.00	28.0	12220
Temp (min) Dec, 19	17.00	1.60	9.20	27.0	12220
Soil condition	1.10	0.61	0.00	2.0	12220
Total Nitrogen (ppm)	759.00	106.00	523.00	1345.0	12220
Total Phosphorus (ppm)	235.00	62.00	142.00	669.0	12220
Extractable Potassium (ppm)	121.00	37.00	56.00	299.0	12220
Water holding capacity (mm)	95.00	22.00	8.90	131.0	12220
Soil pH (depth 0–5cm)	59.00	2.70	54.00	66.0	12220

Notes: This table presents the summary statistics of the control variables in the regression used to estimate household production responses to climate risk.

Table A5: Summary statistics: Mediation analysis

Variable	Mean	Std. Dev.	Min	Max	Obs
Log yield (log kg/ha)	7.40	0.98	2.2	9.3	11459
Fertilizer (kg/ha)	102.00	119.00	0.0	3425.0	11459
Hybrid seed = 1	0.74	0.44	0.0	1.0	11429
Soil condition	1.00	0.57	0.0	2.0	11459
Total Nitrogen (ppm)	758.00	105.00	523.0	1345.0	11459
Total Phosphorus (ppm)	235.00	60.00	142.0	669.0	11459
Extractable Potassium (ppm)	121.00	37.00	56.0	299.0	11459
Water holding capacity (mm)	94.00	22.00	8.9	131.0	11459
Soil pH (depth 0–5cm)	59.00	2.70	54.0	66.0	11459

Notes: This table presents the summary statistics of the plot-level variables used in the mediation analysis. The soil condition variable represents the subjective assessment of the soil quality of the plot, with 0 as low, 1 as medium, and 2 as high.

Table A6: Robustness to using the CV of rainy season rainfall as the risk measure: Diversification

	# of Crop	Gini–Simpson index	# of Plot	Millet	Sorghum
CoV (Nov–Feb)	3.93	0.95*	2.48	0.38	-0.99
	(2.57)	(0.53)	(3.07)	(0.71)	(0.62)
Average precipitation	0.13**	0.033***	0.31***	0.027*	-0.010
	(0.055)	(0.011)	(0.060)	(0.014)	(0.0079)
Prec Nov, 19	-0.073	-0.063	-0.039	-0.059	-0.042
	(0.21)	(0.046)	(0.25)	(0.062)	(0.040)
Prec Dec, 19	0.34**	0.053	0.23	0.11**	0.031
	(0.13)	(0.033)	(0.15)	(0.043)	(0.043)
Prec Jan, 20	-0.31**	-0.031	-0.31**	-0.084*	-0.019
	(0.15)	(0.032)	(0.15)	(0.044)	(0.036)
Prec Feb, 20	-0.27**	-0.038	-0.39***	0.0053	0.00022
	(0.13)	(0.031)	(0.13)	(0.034)	(0.031)
Temp (min) Nov, 19	-0.46***	-0.091***	-0.41***	0.013	0.033
	(0.12)	(0.024)	(0.12)	(0.036)	(0.037)
Temp (max) Nov, 19	0.21**	0.020	0.097	-0.043	-0.016
	(0.094)	(0.020)	(0.11)	(0.026)	(0.017)
Temp (min) Dec, 19	0.27**	0.062**	0.28**	0.023	0.051
	(0.12)	(0.028)	(0.12)	(0.034)	(0.036)
Temp (max) Dec, 19	-0.20	-0.026	-0.33*	0.058	0.052*
	(0.16)	(0.033)	(0.17)	(0.037)	(0.031)
Temp (min) Jan, 20	0.19	0.030	0.023	0.038	-0.0021
	(0.16)	(0.036)	(0.17)	(0.053)	(0.035)
Temp (max) Jan, 20	0.17	0.049	0.25	0.0064	0.057
	(0.18)	(0.038)	(0.20)	(0.057)	(0.044)
Temp (min) Feb, 20	-0.017	0.0060	0.17	-0.052	-0.063**
	(0.14)	(0.029)	(0.13)	(0.053)	(0.027)
Temp (max) Feb, 20	-0.16	-0.045	-0.056	-0.040	-0.096**
	(0.13)	(0.029)	(0.16)	(0.044)	(0.043)
Soil condition = Medium	-0.025	-0.00094	-0.023	-0.0083	-0.0044
	(0.040)	(0.0083)	(0.042)	(0.0053)	(0.0038)
Soil condition = High	0.12**	0.024**	0.17***	0.0081	-0.00071
	(0.051)	(0.010)	(0.051)	(0.0085)	(0.0051)
Total Nitrogen (ppm)	-0.00014	-0.000037	-0.000070	-0.00016	-0.00014
	(0.00052)	(0.00012)	(0.00051)	(0.00011)	(0.000095)
Total Phosphorus (ppm)	-0.0022***	-0.00051***	-0.0024***	-0.00042**	-0.00023
	(0.00062)	(0.00016)	(0.00081)	(0.00020)	(0.00014)
Extractable Potassium (ppm)	0.0010	0.00025	0.0011	0.00089*	0.0012***
	(0.0013)	(0.00029)	(0.0019)	(0.00046)	(0.00037)
Water holding capacity (mm)	0.0018	0.00034	0.0031	0.0011*	0.0011*
	(0.0022)	(0.00056)	(0.0023)	(0.00058)	(0.00061)
Soil pH (depth 0–5cm)	-0.045	-0.0040	0.0021	0.021**	0.0098
	(0.030)	(0.0067)	(0.036)	(0.0080)	(0.0072)
Adj. R-Squared	0.39	0.27	0.39	0.19	0.14
Observations	12220	12220	12220	12220	12220

Notes: Robust standard errors clustered by district are reported in parentheses. Total land size, family size, head's gender, age, years of education, and province fixed effects are included but not reported. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A7: Robustness to using the CV of rainfall between November and February as the risk measure: Diversification

	# of Crop	Gini–Simpson index	# of Plot	Millet	Sorghum
CoV (Nov–Apr)	0.92	0.12	-0.56	0.14	-1.12*
	(2.60)	(0.57)	(3.09)	(0.73)	(0.62)
Average precipitation	0.10*	0.026**	0.28***	0.025*	-0.011
	(0.055)	(0.011)	(0.062)	(0.015)	(0.0074)
Prec Nov, 19	-0.037	-0.054	-0.013	-0.056	-0.048
	(0.21)	(0.046)	(0.25)	(0.062)	(0.039)
Prec Dec, 19	0.36***	0.059*	0.25	0.12**	0.031
	(0.13)	(0.033)	(0.15)	(0.044)	(0.044)
Prec Jan, 20	-0.34**	-0.038	-0.34**	-0.086*	-0.022
	(0.15)	(0.032)	(0.15)	(0.044)	(0.037)
Prec Feb, 20	-0.27**	-0.038	-0.38***	0.0044	0.0089
	(0.13)	(0.031)	(0.13)	(0.035)	(0.033)
Temp (min) Nov, 19	-0.40***	-0.075***	-0.35***	0.018	0.036
	(0.12)	(0.024)	(0.13)	(0.036)	(0.038)
Temp (max) Nov, 19	0.20**	0.018	0.092	-0.043	-0.014
	(0.092)	(0.019)	(0.11)	(0.026)	(0.017)
Temp (min) Dec, 19	0.26**	0.059**	0.27**	0.023	0.048
	(0.12)	(0.029)	(0.12)	(0.035)	(0.036)
Temp (max) Dec, 19	-0.18	-0.022	-0.32*	0.060	0.049
	(0.16)	(0.033)	(0.17)	(0.037)	(0.030)
Temp (min) Jan, 20	0.18	0.028	0.018	0.037	0.00034
	(0.16)	(0.036)	(0.17)	(0.052)	(0.035)
Temp (max) Jan, 20	0.16	0.048	0.25	0.0063	0.056
	(0.18)	(0.039)	(0.20)	(0.058)	(0.045)
Temp (min) Feb, 20	-0.078	-0.011	0.11	-0.057	-0.065**
	(0.13)	(0.028)	(0.12)	(0.051)	(0.028)
Temp (max) Feb, 20	-0.15	-0.043	-0.046	-0.040	-0.095**
	(0.14)	(0.031)	(0.17)	(0.044)	(0.044)
Soil condition = Medium	-0.020	0.00022	-0.019	-0.0079	-0.0050
	(0.040)	(0.0084)	(0.042)	(0.0054)	(0.0037)
Soil condition = High	0.12**	0.025**	0.17***	0.0082	-0.0012
	(0.050)	(0.010)	(0.051)	(0.0086)	(0.0052)
Total Nitrogen (ppm)	-0.00013	-0.000037	-0.000075	-0.00016	-0.00014
	(0.00051)	(0.00012)	(0.00051)	(0.00011)	(0.000096)
Total Phosphorus (ppm)	-0.0023***	-0.00054***	-0.0025***	-0.00043**	-0.00021
	(0.00061)	(0.00015)	(0.00079)	(0.00020)	(0.00014)
Extractable Potassium (ppm)	0.0011	0.00027	0.0011	0.00090*	0.0011***
	(0.0013)	(0.00029)	(0.0019)	(0.00046)	(0.00038)
Water holding capacity (mm)	0.0014	0.00024	0.0027	0.0011*	0.0011*
	(0.0022)	(0.00056)	(0.0023)	(0.00057)	(0.00060)
Soil pH (depth 0–5cm)	-0.041	-0.0028	0.0073	0.021**	0.011
	(0.031)	(0.0068)	(0.037)	(0.0082)	(0.0075)
Adj. R-Squared	0.39	0.27	0.39	0.19	0.14
Observations	12220	12220	12220	12220	12220

Notes: Robust standard errors clustered by district are reported in parentheses. Total land size, family size, head's gender, age, years of education, and province fixed effects are included but not reported. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A8: Robustness to using the CV of rainy season rainfall as the risk measure: Investments in fertilizer and seeds

	Basal (/ha)	Top (/ha)	Total (/ha)	Hybrid seed
CoV (Nov–Feb)	-47.2	22.8	-24.4	2.39***
	(91.7)	(96.6)	(186.3)	(0.69)
Average precipitation	8.36***	10.2***	18.5***	0.033*
	(2.09)	(2.18)	(4.21)	(0.018)
Prec Nov, 19	-13.6**	-11.6*	-25.2**	0.027
	(6.26)	(6.42)	(12.6)	(0.065)
Prec Dec, 19	15.2***	15.4***	30.6***	0.12***
	(5.49)	(5.45)	(10.9)	(0.038)
Prec Jan, 20	2.50	0.047	2.54	0.088*
	(6.20)	(6.25)	(12.4)	(0.048)
Prec Feb, 20	-20.2***	-23.1***	-43.3***	-0.17***
	(4.03)	(3.81)	(7.58)	(0.035)
Temp (min) Nov, 19	1.03	0.29	1.33	-0.018
	(4.00)	(4.22)	(8.12)	(0.028)
Temp (max) Nov, 19	-0.55	0.70	0.15	0.034
	(3.46)	(3.70)	(7.12)	(0.026)
Temp (min) Dec, 19	2.04	0.62	2.66	0.019
	(5.30)	(5.32)	(10.5)	(0.039)
Temp (max) Dec, 19	-0.52	-1.26	-1.78	-0.015
	(6.20)	(6.45)	(12.6)	(0.060)
Temp (min) Jan, 20	-16.7**	-14.4**	-31.1**	-0.051
	(6.62)	(6.96)	(13.5)	(0.056)
Temp (max) Jan, 20	1.11	0.91	2.02	0.027
	(6.05)	(5.97)	(11.9)	(0.050)
Temp (min) Feb, 20	10.2*	10.9*	21.1*	-0.0048
	(5.74)	(5.85)	(11.5)	(0.052)
Temp (max) Feb, 20	-1.37	-2.71	-4.08	-0.038
	(6.55)	(6.59)	(13.1)	(0.058)
Soil condition = Medium	6.85***	7.32***	14.2***	0.065***
	(1.77)	(1.56)	(3.28)	(0.017)
Soil condition = High	5.57***	6.35***	11.9***	0.079***
	(1.97)	(1.87)	(3.77)	(0.020)
Total Nitrogen (ppm)	0.034*	0.028	0.062	0.00028
	(0.020)	(0.021)	(0.040)	(0.00017)
Total Phosphorus (ppm)	-0.11***	-0.096***	-0.20***	-0.00038*
	(0.026)	(0.028)	(0.053)	(0.00022)
Extractable Potassium (ppm)	0.025	0.021	0.045	0.00055
	(0.056)	(0.061)	(0.12)	(0.00047)
Water holding capacity (mm)	0.11	0.12	0.23	0.0021***
	(0.089)	(0.091)	(0.18)	(0.00078)
Soil pH (depth 0–5cm)	-0.18	-0.38	-0.56	-0.00050
	(1.28)	(1.30)	(2.57)	(0.0087)
Adj. R-Squared	0.26	0.26	0.27	0.24
Observations	12220	12220	12220	12220

Notes: Robust standard errors clustered by district are reported in parentheses. Total land size, family size, head's gender, age, years of education, and province fixed effects are included but not reported. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A9: Robustness to using the CV of rainfall between November and February as the risk measure: Investments in fertilizer and seeds

	Basal (/ha)	Top (/ha)	Total (/ha)	Hybrid seed
CoV (Nov–Apr)	-128.3	-64.0	-192.2	2.13***
	(85.3)	(89.9)	(173.3)	(0.64)
Average precipitation	7.73***	9.44***	17.2***	0.030
	(2.14)	(2.17)	(4.25)	(0.018)
Prec Nov, 19	-13.6**	-11.2*	-24.9**	0.043
	(6.16)	(6.29)	(12.3)	(0.065)
Prec Dec, 19	15.6***	15.9***	31.5***	0.13***
	(5.48)	(5.48)	(10.9)	(0.040)
Prec Jan, 20	1.60	-0.85	0.75	0.088*
	(6.25)	(6.33)	(12.5)	(0.050)
Prec Feb, 20	-19.1***	-22.5***	-41.7***	-0.18***
	(4.01)	(3.79)	(7.57)	(0.036)
Temp (min) Nov, 19	2.72	2.06	4.78	-0.013
	(3.93)	(4.20)	(8.04)	(0.029)
Temp (max) Nov, 19	-0.47	0.64	0.17	0.029
	(3.45)	(3.69)	(7.10)	(0.026)
Temp (min) Dec, 19	1.41	0.11	1.52	0.023
	(5.47)	(5.52)	(10.9)	(0.039)
Temp (max) Dec, 19	-0.58	-1.08	-1.66	-0.0074
	(6.44)	(6.67)	(13.1)	(0.058)
Temp (min) Jan, 20	-16.6**	-14.4**	-31.0**	-0.057
	(6.49)	(6.87)	(13.3)	(0.056)
Temp (max) Jan, 20	0.99	0.81	1.80	0.028
	(6.08)	(6.02)	(12.0)	(0.051)
Temp (min) Feb, 20	8.63	9.19	17.8	-0.011
	(5.51)	(5.72)	(11.1)	(0.050)
Temp (max) Feb, 20	-1.01	-2.39	-3.40	-0.039
	(6.57)	(6.58)	(13.1)	(0.058)
Soil condition = Medium	6.86***	7.39***	14.3***	0.067***
	(1.77)	(1.56)	(3.27)	(0.017)
Soil condition = High	5.52***	6.34***	11.9***	0.080***
	(1.94)	(1.84)	(3.72)	(0.020)
Total Nitrogen (ppm)	0.034*	0.027	0.061	0.00029*
	(0.019)	(0.021)	(0.040)	(0.00017)
Total Phosphorus (ppm)	-0.11***	-0.097***	-0.20***	-0.00044**
	(0.026)	(0.028)	(0.054)	(0.00022)
Extractable Potassium (ppm)	0.021	0.019	0.040	0.00064
	(0.055)	(0.059)	(0.11)	(0.00046)
Water holding capacity (mm)	0.10	0.11	0.21	0.0020***
	(0.088)	(0.090)	(0.18)	(0.00077)
Soil pH (depth 0–5cm)	0.048	-0.20	-0.15	-0.0017
	(1.29)	(1.33)	(2.61)	(0.0089)
Adj. R-Squared	0.26	0.26	0.27	0.24
Observations	12220	12220	12220	12220

Notes: Robust standard errors clustered by district are reported in parentheses. Total land size, family size, head's gender, age, years of education, and province fixed effects are included but not reported. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A10: Robustness to using the CV of rainy season rainfall as the risk measure: Family labor and land utilization

	Labor	Female labor	Male labor	Area planted/Area field
CoV (Nov–Feb)	0.85	-0.86	1.71*	0.54
	(1.31)	(0.88)	(0.94)	(0.45)
Average precipitation	0.035	0.019	0.017	-0.034***
	(0.036)	(0.024)	(0.021)	(0.0089)
Prec Nov, 19	0.31*	0.10	0.20**	-0.0026
	(0.16)	(0.095)	(0.098)	(0.029)
Prec Dec, 19	-0.073	-0.041	-0.032	0.035*
	(0.097)	(0.060)	(0.059)	(0.021)
Prec Jan, 20	0.15	0.13**	0.024	-0.012
	(0.092)	(0.057)	(0.061)	(0.021)
Prec Feb, 20	-0.16*	-0.053	-0.10*	0.023
	(0.086)	(0.057)	(0.056)	(0.020)
Temp (min) Nov, 19	-0.062	0.014	-0.075	-0.023
	(0.075)	(0.040)	(0.054)	(0.019)
Temp (max) Nov, 19	0.100*	-0.014	0.11***	0.049***
	(0.056)	(0.037)	(0.035)	(0.016)
Temp (min) Dec, 19	-0.012	-0.057	0.045	0.0058
	(0.084)	(0.053)	(0.052)	(0.021)
Temp (max) Dec, 19	-0.20**	-0.059	-0.14**	0.0030
	(0.090)	(0.061)	(0.054)	(0.023)
Temp (min) Jan, 20	-0.30***	-0.058	-0.24***	0.063**
	(0.11)	(0.062)	(0.070)	(0.030)
Temp (max) Jan, 20	-0.085	-0.031	-0.054	-0.029
	(0.13)	(0.070)	(0.081)	(0.025)
Temp (min) Feb, 20	0.37***	0.077	0.29***	-0.067***
	(0.11)	(0.061)	(0.072)	(0.022)
Temp (max) Feb, 20	0.18	0.11	0.073	-0.014
	(0.12)	(0.065)	(0.077)	(0.022)
Soil condition = Medium	0.011	-0.0023	0.013	0.0023
	(0.047)	(0.026)	(0.031)	(0.0047)
Soil condition = High	-0.0040	0.015	-0.019	-0.0034
	(0.055)	(0.031)	(0.036)	(0.0065)
Total Nitrogen (ppm)	-0.00028	-0.00016	-0.00012	-0.000024
	(0.00042)	(0.00026)	(0.00022)	(0.00011)
Total Phosphorus (ppm)	0.00064	0.00011	0.00053*	-0.00010
	(0.00057)	(0.00034)	(0.00030)	(0.00019)
Extractable Potassium (ppm)	0.00030	0.00084	-0.00053	0.00029
	(0.0011)	(0.00078)	(0.00069)	(0.00035)
Water holding capacity (mm)	-0.0017	-0.0010	-0.00066	0.000081
	(0.0018)	(0.0011)	(0.0012)	(0.00041)
Soil pH (depth 0–5cm)	-0.012	0.0060	-0.018	-0.0063
	(0.019)	(0.012)	(0.012)	(0.0050)
Adj. R-Squared	0.54	0.37	0.38	0.43
Observations	12220	12220	12220	12220

Notes: Robust standard errors clustered by district are reported in parentheses. Total land size, family size, head's gender, age, years of education, and province fixed effects are included but not reported. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A11: Robustness to using the CV of rainfall between November and February as the risk measure: Family labor and land utilization

	Labor	Female labor	Male labor	Area planted/Area field
CoV (Nov–Apr)	0.54	-0.43	0.97	0.62
	(1.36)	(0.86)	(0.88)	(0.53)
Average precipitation	0.032	0.023	0.0096	-0.034***
	(0.036)	(0.025)	(0.020)	(0.0087)
Prec Nov, 19	0.31*	0.094	0.22**	0.00068
	(0.16)	(0.096)	(0.096)	(0.030)
Prec Dec, 19	-0.070	-0.045	-0.025	0.036*
	(0.098)	(0.060)	(0.060)	(0.021)
Prec Jan, 20	0.15	0.13**	0.018	-0.010
	(0.092)	(0.059)	(0.061)	(0.021)
Prec Feb, 20	-0.16*	-0.050	-0.11*	0.019
	(0.084)	(0.056)	(0.056)	(0.021)
Temp (min) Nov, 19	-0.056	0.0055	-0.061	-0.025
	(0.077)	(0.040)	(0.056)	(0.019)
Temp (max) Nov, 19	0.098*	-0.013	0.11***	0.048***
	(0.056)	(0.037)	(0.036)	(0.016)
Temp (min) Dec, 19	-0.012	-0.056	0.044	0.0076
	(0.083)	(0.053)	(0.052)	(0.021)
Temp (max) Dec, 19	-0.20**	-0.062	-0.13**	0.0046
	(0.091)	(0.060)	(0.054)	(0.022)
Temp (min) Jan, 20	-0.30***	-0.056	-0.25***	0.061**
	(0.10)	(0.063)	(0.069)	(0.031)
Temp (max) Jan, 20	-0.085	-0.031	-0.055	-0.029
	(0.13)	(0.070)	(0.083)	(0.025)
Temp (min) Feb, 20	0.36***	0.086	0.28***	-0.066***
	(0.12)	(0.059)	(0.078)	(0.023)
Temp (max) Feb, 20	0.18	0.11	0.075	-0.014
	(0.12)	(0.065)	(0.079)	(0.022)
Soil condition = Medium	0.012	-0.0032	0.015	0.0026
	(0.047)	(0.026)	(0.031)	(0.0048)
Soil condition = High	-0.0037	0.015	-0.018	-0.0032
	(0.055)	(0.031)	(0.036)	(0.0064)
Total Nitrogen (ppm)	-0.00028	-0.00016	-0.00012	-0.000021
	(0.00042)	(0.00026)	(0.00023)	(0.00011)
Total Phosphorus (ppm)	0.00062	0.00013	0.00049	-0.00011
	(0.00056)	(0.00034)	(0.00030)	(0.00019)
Extractable Potassium (ppm)	0.00033	0.00081	-0.00048	0.00031
	(0.0011)	(0.00078)	(0.00069)	(0.00036)
Water holding capacity (mm)	-0.0017	-0.00099	-0.00075	0.000085
	(0.0018)	(0.0011)	(0.0012)	(0.00041)
Soil pH (depth 0–5cm)	-0.012	0.0056	-0.018	-0.0070
	(0.019)	(0.012)	(0.012)	(0.0049)
Adj. R-Squared	0.54	0.37	0.38	0.43
Observations	12220	12220	12220	12220

Notes: Robust standard errors clustered by district are reported in parentheses. Total land size, family size, head's gender, age, years of education, and province fixed effects are included but not reported. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

A.3 Identification of ACDE

The ACDE can be identified under these two assumptions.

$$
\begin{aligned}
\gamma(r, m, x) &= E[Y_i(r, m) - Y_i(r, m')|X_i] \\
&= E[Y_i(r, m)|X_i, Z_i, R_i] - E[Y_i(r, m')|X_i, Z_i, R_i] \\
E[Y_i - \gamma(r, m, x)|R_i, X_i] &= E[Y_i(r, m')|X_i] \\
E[Y_i - \gamma(r, M_i, x)|R_i, X_i] &- E[Y_i - \gamma(r', M_i, x)|R_i, X_i] \\
&= E[Y_i(r, M_i)|X_i] - E[Y_i(r', M_i)|X_i] \\
&= ACDE(x)
\end{aligned}
$$

A.4 Bias Form

$$
M_l = \gamma_0 + \gamma_1 CV_w^{12} + X_l'\gamma_2 + Z_l'\gamma_3 + \delta_p + \xi_l \tag{7}
$$

Acharya et al. (2016) demonstrate that the bias of the estimator of the ACDE is:

$$
\begin{aligned}
&plim\ \widehat{ACDE} - ACDE \\
&= -\tilde{\delta}\frac{\tilde{\delta}_y}{\delta_m}\sqrt[2]{(1-\tilde{\rho}^2)/(1-\rho^2)},
\end{aligned}
$$

where ρ is the correlation coefficient between the error terms of equation (5) and equation (7) and δ is the effect of the treatment on the mediator.